th Annual South-East European Doctoral Student Conference

Infusing Research and Knowledge in South-East Europe

Edited by: K. Bratanis, D. Dranidis, P. Koktsidis, L. Lazouras, E. Nikolaidou



Proceedings of the 7th Annual South-East European Doctoral Student Conference

Edited by

K. Bratanis, D. Dranidis, P. Koktsidis, L. Lazouras, E. Nikolaidou

ISBN 978-960-9416-05-4 ISSN 1791-3578

©SOUTH-EAST EUROPEAN RESEARCH CENTRE, 2012

24 P. Koromila str. 54622 Thessaloniki, Greece Tel: +30 2310 253 477 Fax: +30 2310 234 205 http://www.seerc.org

Preface

These proceedings represent the work of contributions to the seventh (7th) Annual SEERC Doctoral Student Conference (DSC2012) hosted by The University of Sheffield International Faculty, CITY College and organised by the South East European Research Centre - Thessaloniki, Greece.

DSC 2012 has grown and continues to evolve. Now is the seventh year the key aim remains to provide an opportunity for PhD students to receive advice from experts in their chosen field of research. Having identified academic isolation as a problem that many doctoral students face today, SEERC aims to bring researchers together for establishing collaborative links between disciplines, for testing the ground for innovative ideas and for engaging the wider academic community.

Building on the success of the past six conferences, this years conference attracted a large number of submissions resulting in 106 presentations of full papers. The audience of the conference expanded beyond the boundaries of South East Europe confirming the need for Doctoral Students to come together, discuss their experiences and gain external feedback to their work as well as listen to the progress and methodology of fellow PhD candidates.

These papers represent research from Albania, Algeria, Australia, Austria, Bosnia and Herzegovina, Bangladesh, Bulgaria, China, Czech Republic, Cyprus, Egypt, Finland, France, F.Y.R Macedonia, Georgia, Germany, Ghana, Greece, Hungary, Italy, Morocco, Poland, Romania, Serbia, Slovakia, Slovenia, South Africa, Sweden, Switzerland, Turkey, United Kingdom.

September 2012

Prof. Panayiotis H. Ketikidis Chairperson of DSC2012

Conference Organisation

DSC2012 is organized by the South-East European Research Centre (SEERC), an overseas research centre of The University of Sheffield, established as a non-profit legal entity in Thessaloniki, Greece. SEERC was founded by CITY College, the University's Intentional Faculty, in 2003.

Organising Committee

Ms. Andreea Teodora Balan, South-East European Research Centre
Mr. Konstantinos Bratanis, South-East European Research Centre
Dr. Dimitris Dranidis, The University of Sheffield International Faculty, CITY College
Ms. Athanasia Halatzouka, South-East European Research Centre
Dr. Pavlos Koktsidis, The University of Sheffield International Faculty, CITY College
Dr. Lambros Lazouras, The University of Sheffield International Faculty, CITY College
Dr. Eftychia Nikolaidou, The University of Sheffield International Faculty, CITY College
Ms. Danai Triantafylli, South-East European Research Centre

Scientific Committee Chairs

Enterprise, Innovation and Development

Dr. Eftychia Nikolaidou, The University of Sheffield International Faculty, CITY College Dr. Tim Vorley, The University of Sheffield

Information and Communication Technologies

Dr. Dimitris Dranidis, The University of Sheffield International Faculty, CITY College Dr. Anthony Simons, The University of Sheffield

Society and Human Development: Psychology, Politics, Sociology, and Education

Dr. Lambros Lazouras, The University of Sheffield International Faculty, CITY College Dr. Pavlos Koktsidis, The University of Sheffield International Faculty, CITY College

Scientific Committee

Enterprise, Innovation and Development

Dr. Kostas Alexandris, Aristotle University of Thessaloniki Dr. Dimitris Dadakas, The University of Sheffield International Faculty, CITY College Dr. Anna Watson, London Fashion College, University of London Dr. Howard Viney, London Business School and Open University Prof. Paul Dunne, University of Cape Town Dr. Ian Roper, Middlesex University Business School Dr. Alexandros Kapoulas, The University of Sheffield International Faculty, CITY College Mr. Fotis Missopoulos, The University of Sheffield International Faculty, CITY College Dr. Nikolaos Mylonidis, University of Ioannina Dr. Eftychia Nikolaidou, The University of Sheffield International Faculty, CITY College Dr. Leslie T. Szamosi, The University of Sheffield International Faculty, CITY College Dr. Athina Vasilaki, The University of Sheffield International Faculty, CITY College Prof. Geoff Wood, The University of Sheffield Dr. Konstantinos Priporas, University of Macedonia Dr. Tim Vorley, The University of Sheffield Prof. Christos Kollias, University of Thessaly Dr. Alexandros Psychogios, The University of Sheffield International Faculty, CITY College Dr. Dimitris Georganakos, Goethe University Dr. Alexandros Zangelidis, University of Aberdeen Dr. Panos Piperopoulos, University of Newcastle

Information and Communication Technologies

Dr. Galia Angelova, Bulgarian Academy of Sciences
Dr. Costin Badica, University of Craiova
Prof. Tudor Balanescu, University of Pitesti
Prof. Omer Faruk Bay, Gazi University
Prof. Dimitar Birov, University of Sofia
Assist. Prof. Alexander Chatzigeorgiou, University of Macedonia
Prof. Betim Çiço, Polytechnic University of Tirana
Dr. Tony Cowling, The University of Sheffield
Dr. Konstantinos Dimopoulos, The University of Sheffield International Faculty, CITY College
Prof. Christos Douligeris, University of Piraeus Dr. Dimitris Dranidis, The University of Sheffield International Faculty, CITY College Dr. George Eleftherakis, The University of Sheffield International Faculty, CITY College Prof. Neki Frasheri, Polytechnic University of Tirana Assoc. Prof. Akli Fundo, Polytechnic University of Tirana Dr. Christos Georgiadis, University of Macedonia Prof. Florin Gorunescu, University of Craiova Mr. Thanos Hatziapostolou, The University of Sheffield International Faculty, CITY College Dr. Stylianos Hatzipanagos, King's College London Prof. Florentin Ipate, University of Pitesti Prof. Petros Kefalas, The University of Sheffield International Faculty, CITY College Prof. Yannis Manolopoulos, Aristotle University of Thessaloniki Assoc. Prof. Krasimir Manev, University of Sofia Prof. Roger Moore, The University of Sheffield Dr. Miguel Nunes, The University of Sheffield Dr. Iraklis Paraskakis, South-East European Research Centre Assoc. Prof. Anton Petkov, University of Sofia Dr. Ilias Sakellariou, University of Macedonia Prof. Demetrios Sampson, University of Piraeus Dr. Anthony Simons, The University of Sheffield Mrs. Anna Sotiriadou, The University of Sheffield International Faculty, CITY College Dr. Ioanna Stamatopoulou, The University of Sheffield International Faculty, CITY College Dr. Ana Vasconcelos, The University of Sheffield Dr. Anca Vasilescu, University of Brasov Dr. Husnu Yenigun, Sabanci University of Turkey Prof. Katerina Zdravkova, University of Ss Cyril & Methodius

Society and Human Development: Psychology, Politics, Sociology, and Education

Dr. Nikolaos Tsigilis, Aristotle University of Thessaloniki

Dr. Tom Stafford, The University of Sheffield

Dr. Chris Jones, The University of Sheffield

Dr. Elvira Masoura, Aristotle University of Thessaloniki

Dr. Mary Kosmidis, Aristotle University of Thessaloniki

Dr. Antonia Ypsilanti, Aristotle University of Thessaloniki

Dr. Angeles F Estevez, Universidad de Almeria

Dr. Kwang-Hyuk Lee, The University of Sheffield

Dr. Panagiotis Bamidis, Aristotle University of Thessaloniki

Dr. Aggelos Rodafinos, The University of Sheffield International Faculty, CITY College

Dr. Suzie Savvidou, The University of Sheffield International Faculty, CITY College Dr. Martha Kelpi, The University of Sheffield International Faculty, CITY College Dr. Lazuras Lambros, The University of Sheffield International Faculty, CITY College Dr. Tatiana Tairi, The University of Sheffield International Faculty, CITY College Dr. Efrosini Kalyva, The University of Sheffield International Faculty, CITY College Ms. Agnes Brunet, The University of Sheffield International Faculty, CITY College Prof. Simon Bulmer, The University of Sheffield Prof. Bob Deacon, The University of Sheffield Dr. Sara Hannam, The University of Sheffield International Faculty, CITY Colleae Dr. Nicola Rehling, Aristotle University of Thessaloniki Prof. Nicos Sifakis, Hellenic Open University Dr. Paul Stubbs, Institute of Economics Zagreb Dr. Natalia Vlas, Babes-Bolyai University Prof. Vasile Boari, Babes-Bolyai University Prof. Theodore Chadjipandelis, Aristotle University of Thessaloniki Dr. Christos Frangonikolopoulos, Aristotle University of Thessaloniki Prof. Ralph Negrine, The University of Sheffield Dr. Pavlos Kotsidis, The University of Sheffield International Faculty, CITY College Dr. Filipos Proedrou, The University of Sheffield International Faculty, CITY College Dr. Maria Xenitidou, University of Surrey Dr. Zoi Tatsioka, The University of Sheffield International Faculty, CITY College Prof. Paul Knepper, The University of Sheffield

Sponsoring Institutions

South-East European Research Centre (SEERC), Greece The University of Sheffield International Faculty, CITY College, Greece The University of Sheffield, United Kingdom Urenio Research Unit, Aristotle University, Greece VUZF University, Bulgaria University of Sofia, Bulgaria Polytechnic University of Tirana, Albania

Table of Contents

Enterprise, Innovation and Development

Discretionary Fiscal Policy in the EMU Context: An EmpiricalApproach (1981-2010) & the Recent Reform of European GovernanceDionysios K. Solomos, Dimitrios N. Koumparoulis, and DimitriosP. Kamsaris	19
Fiscal Decentralization in Rural Local Governments in Mexico: Changes in Accountability and Entrepreneurship in the Local Government Structures	45
Foreign Direct Investment through the Imitation Procedure: The Case of China	62
The Contribution of EU Regional Policy to Structural Change and to Adaptation to Globalization in less Developed Regions	77
Conspicuous Consumption: Essential Concepts Armando Memushi	91
Market Activity Co-movements: Evidence from the S&P 500 Index, Crude Oil and Gold Futures Market <i>Michael Soucek</i>	105
Factors Influencing Bank Profitability Comparison of Banks in Serbia, Croatia, Slovenia and the 10 Largest Banks in EU Area Ana Polovina and Ljiljana Popovic	120
Empirical Investigation of Credit Risk Determinants: The Case of Tajikistan Banking Industry Mukhayo Khodzhibaeva and Eftychia Nikolaidou	125
US Corporate Bond Yield Spread: A Default Risk Debate Noaman Shah	138
Predicting Stock Returns: Evidence from the Istanbul Stock Exchange $G\ddot{u}nes\ Topçu$	153
Crisis on the Hungarian Government Bond Markets: is that a Liquidity Problem?	167

Privatization and Market Liquidity in Croatia and Serbia Goran Zafirov	177
Terrorism and Capital Markets: The Effects of the Istanbul Bombings Nikos Christofis, Christos Kollias, Stephanos Papadamou, and Apostolos Stagiannis	189
Investigating the Causality between Financial Development and Economic Growth in the Emerging Economies of Europe, Specifically: Albania and Turkey Esida Gila Gourgoura	202
Benchmarking Total Factor Productivity in the Regulation of Natural Gas Distribution	208
Lean Six Sigma in Food Industry in Time of Globalization Ivana Nikolic	213
Exploring the Potential of a Satisfied and Loyal Employee as Effective and Best Marketing and/or PR Agent of his/her Company Lana Milošević	228
Understanding and Improving Employee Engagement in Organizations An Empirical Study in Indonesia	253
Public Human Resources Management Role: Ways of Implementing Organizational Social Responsibility Strategy Straia (Ilieş) Camelia Olivia	261
Social Enterprise and Employee Empowerment: The Role of Employee Ownership	274
Determinants of the HRM Policies and Practices in Hotel MNEs: the View from the HR Department <i>Giovanni Oscar Serafini</i>	279
Understanding Training in a Complexity Theory Context: Training Process as a Simplifier of Organizational Complexity Cristina Catalina Chirita and Alexandros Psychogios	304
Exploring the Relationship between the Learning Organization and Business Performance Bojana Milić	318

High Performance Work Systems and Organizational Performance in Multinational Subsidiaries Operating in the Balkan Region Andreea Teodora Balan, Alexandros Psychogios, and Athina Vasilaki	326
Business Relationships between Non- and For- profit Sectors Maria Magyar	343
Keys Success Factors of New Service Development from the Open Innovation Perspective: Benefits and Implications Eriona Shtembari and Gert Pustina	353
Determining the Impact of Interorganizational Relationships (IOR) on Innovation: Theoretical Framework Building Danimir Ljepava	375
Strategic Planning Practices in The State Owned Banks The Indonesian Case	388
Role Of Leadership In Organizational Effectiveness In Albania Dritan Shoraj and Dorina Zenelaj	407
The Impact of Informality at the Labour Market: The case of Construction Industry during Albanian Transition Period Enida Pulaj	422
The Influence of Gender and Location on Young Saudi Arabian Customers Choice, Satisfaction and Usage of Mobile Phone Services Badr Alharbi	430
What is Business Failure? A Philosophical Perspective Joni Salminen	438
Information and Communication Technologies	
An Engineering Perspective on Emergence Konstantinos Rousis, George Eleftherakis, and Anthony J. Cowling	453
Stability Properties of Stochastic Neural Networks Biljana Tojtovska	464
Tools for Visual Simulation of MAS Models Isidora Petreska	475
A Thermographic Atlas of the Human Body Viktoria Rajtukova, Monika Michalikova, Teodor Toth, and Jozef Zivcak	481
A Three-Dimensional Cellular Automaton for Network Traffic Simulation	492

A Three-Dimensional Cellular Automaton for Network Traffic Simulation 492 Stylianos Dimitriou

Open Data Portal based on Semantic Web Technologies Martin Kostovski, Milos Jovanovik, and Dimitar Trajanov	504
Enhancing Scholarly Communication Results and Search Experience by Interlacing Relevant Scientific Repositories Arben Hajra, Vladimir Radevski, and Atif Latif	517
Relational Databases in Data Mining on Grid <i>Ğorgi Kakaevski and Anastas Mišev</i>	528
An Ontology-driven Approach to Self-management in Cloud Application Platforms Rustem Dautov, Iraklis Paraskakis, and Dimitrios Kourtesis	539
Cloud Computing and E-Learning: a Review Krenare Pireva, Petros Kefalas, and Tony Cowling	551
Addressing the Challenge of Application Portability in Cloud Platforms Fotis Gonidis, Iraklis Paraskakis, and Dimitrios Kourtesis	565
Security Assessment of OpenStack Open Source Cloud Solution Sasko Ristov, Marjan Gusev, and Magdalena Kostoska	577
Fleet Management Solution based on Cloud Computing: Improving Availability of Additional Services Dashmir Istrefi and Betim Çiço	588
Load Balancing in Distributed Systems and Ad-hoc Networks using a Fairness-oriented Approach Stylianos Dimitriou and Elias Christodoulou	598
A New Public Key Cryptosystem with Key Escrow Capabilities	610
A Secure Authenticated Group Key Transfer Protocol Adrian Atanasiu and Adela Georgescu	625
E-Voting algorithm dilemma: Blind Signatures or Homomorphic Encryptions Pance Ribarski and Ljupcho Antovski	635
DSP Tools for Signal Measurement Performance Analysis	645

Koffi A. DOTCHE

Society and Human Development Psychology, Politcs, Sociology and Education

Violent Crime in Romania Estimation, Etiology and Social Prevention Simona Ionela Mihaiu	657
The Protection of Vulnerable Groups in Albania	670
Womens Informal Employment as an Engine of Poverty Reduction Souad Belhorma	685
Power relations between women and men in Roma communities Andreea Bragă	689
A Survey of Intention to Leave among Nurses in the Hail Regions Hospitals in Saudi Arabia: The Role of Job Satisfaction, Job Stress and Burnout	694
Leadership Styles of Nurse Managers at The Ministry of Health Hospitals of Saudi Arabia Farhan Alshammari	699
Nurses Knowledge Regarding Pain Management in Hail Region Hospitals Saudi Arabia	704
Deception and its Possible Relation to Emotional Intelligence among Respondents in the Czech Republic Lenka Mynařková	710
The Construction and Evolution of the Romantic Relationships in Cyberspace	725
Heart Rate Variability During Decision-Making in Children Considering Piagets Theory of Cognitive Development Mária Túnyiová and Ivan Sarmány-Schuller	730
Substance Misuse Treatment Process Evaluation: A Dimensional Trait-based Conceptualization <i>Fivos E. Papamalis, Efrosini Kalyva, and Petra S. Meier</i>	735
Is Attention Influenced by Bilingualism in Young Children of Low Socioeconomic Status? Aristea I. Ladas, Ana B. Vivas and Michael Siegal	747

XIV

Risky Decision-Making of Slovak Undergraduate Students With and Without Migration Experience <i>Eva Drobná and Viera Bačová</i>	764
Self-affirmation and Doping Behaviour Vassilis Barkoukis, Lambros Lazuras, and Peter Harris	772
Can Drivers Cognitive Biases and Driving Behaviour be Modified by Self-affirmation Methods? Danai Triantafylli	776
Optimisation of Solar Thermal Panel Use	784
Daily Diary Study of Marital Support in Couples Dealing with Cancer Aleksandra Gronostaj	793
Better Life Indicators with Principal Component Analysis Vasileios Ismyrlis	800
An Econometric and Social Analysis of the Impact of Tobacco Control Policies in the Hospitality and Tourism Industry in Cyprus Christos Savva, Chrystalla Kapetaniou, Elpidoforos Soteriades, and Lambros Lazouras	817
Motivation of Students at the Palack´ University in Olomouc for Their Specialization in Teacher Profession Pavla Nádvorníková	823
Strategic and Doctrinal Framework of Western Balkans Countries Granit Zela	831
Rearranging the puzzles of security in the Western Balkans: a comparative study of intra-state armed groups	850
Between National Solidarity and Local Interests The Pro-governmental Political Orientation of the Romanians in Hungary (the End of the 19th Century the Beginning of the 20th Century) <i>Iudean Ovidiu Emil</i>	858
The Romanian Administrative Elite from Transylvania in the Liberal Period (1861-1867) Alexandru Onojescu	873
Political and Diplomatic Relations Between Romania and the Ottoman Empire (1878-1900) Cecilia Alina Sava	888

Governmental and Non-governmental Action Against Organized Crime and Drug Trafficking in Serbia Darja Koturovic	896
Party Structure and Organization in Post-Communist Albania: Explaining Centre-Right Party Success Anjeza Xhaferaj	903
Corporate Social Responsibility: A Comparison Between Large and Medium Companies in UK Maryam Sadeghi	916
Neo-Gramscianism and European Integration Mihail Caradaică	923
Anti-Americanism In Europe: A Cultural Problem Florenca Korbi	935
Representations of the Iraq War in Two Doctrines: Neoconservatism and Interventionism Oana Elena Brânda	948
Jewish Converts, Their Communities and Experiences of Social Inclusion and Exclusion in Post-apartheid South Africa <i>Elli Kriel</i>	963
Neobohemia: a Dispute over New Artistic Spaces in a Postmodern City $Weronika\ Maćków$	980
Albanian Language under the Pressure of Social Development Adelina Sula and Shezai Rrokaj	987
Washback of State Matura Examination on Teaching Jaup Zenuni	994
Efficient Breach of the Contract: an Interdisciplinary Approach (Legal and Economic) Delia-Ionela Ropan	1013
Introduction to the Study of Public Transport Accessibility	1025
Author Index	1037
E-Papers	1041

 $\mathbf{X}\mathbf{V}\mathbf{I}$

Enterprise, Innovation and Development

Discretionary Fiscal Policy in the EMU Context: An empirical approach (1981-2010) & the Recent Reform of European Governance

Dionysios K. Solomos¹, Dimitrios N. Koumparoulis², Dimitrios P. Kamsaris³

 ¹ Ph.D candidate, Department of Economics, UGSM-Monarch Business School Switzerland, dionysios.solomos@ugsm-monarch.com
 ² Department of Economics, UGSM-Monarch Business School Switzerland, dr.koumparoulis@ugsm-monarch.com
 ³ Department of Economics, UGSM-Monarch Business School Switzerland, dr.kamsaris@ugsm-monarch.com

Abstract. The paper attempts to investigate how the Maastricht criteria and the Stability and Growth Pact have impaired the capacity of EZ national authorities to conduct discretionary fiscal policy. We estimate fiscal determinants for the structural (discretionary) fiscal deficit over the period of 1981-2010, estimating panel data equations in order to increase the strength of the test by enhancing the time series dimension of the data by the cross section. We find that the degree of the countrer-cyclicality of discretionary fiscal policy has been reduced significantly after the Maastricht Treaty. Also, there is empirical evidence that national fiscal rules have a significant positive impact in budgetary outcomes. Regarding the recent reform of the European governance framework, we consider that the context of the reform seems incapable of dealing with the factors which are responsible for the sovereign debt crisis. We stress the need for reforms in the financial sector which seem necessary to ensure in association with sound fiscal policies the stability in euro area.

Keywords: *discretionarity, Stability and Growth Pact, national fiscal rules, sustainability, fiscal discipline.*

1. Introduction

The debt crisis in the Eurozone (EZ) is one of the most crucial challenges in modern economic history. The decline of public finances of many member states constitutes an explosive mixture that threatens the existence of the Eurozone itself and has significant impacts for the global economy. As the leadership of EU is still searching for solutions and actions which will counteract the problem and restore the stability, the analysis of the causes which led to the systemic crisis, the examination of the institutional framework of European governance and the evaluation of the measures and policies adopted, are considered as matters of great importance which require extensive analysis and critical approach. The current paper attempts to approach these aspects from the perspective of fiscal policy, examining them from both a theoretical and empirical background.

Membership in the Eurozone implies that apart from the delegation of monetary policy to the conservative European Central Bank (ECB), there are obligations and restrictions on fiscal policy for the member states. Even though the latter remains a political tool into the hands of national authorities, they must comply with the rules set at EU level in order the stability to be achieved. The implementation of sound and disciplined fiscal policies and the coordination of them in central level are key elements for ensuring the financial stability in the eurosystem.

As the project of European integration was not associated with a fiscal union, the member states have maintained the independence of their national fiscal policies. However, their autonomy is limited significantly by the existence of the euro zone's supranational fiscal rule. The Stability and Growth Pact (SGP) offers guidance and coordinates the implementation of fiscal policy through the obligation of member states to respect the rules and the procedures that the Pact predicts. The target of the SGP is to perform effectively a dual function: to ensure the sustainability of public finances prompting the governments to run sound policies and to provide the necessary flexibility so as the national authorities to be able to deal with the negative fluctuations of the business cycle.

This paper consists of two parts. In the first section, we are trying to acknowledge the particular structure of fiscal policy in EMU context. We examine empirically how the restrictions of Maastricht criteria and the SGP have impaired the capacity of national authorities to run countercyclical discretionary fiscal policies and whether the stabilization function of fiscal policy is associated with less procyclical discretionary policies. The division of the EZ countries into two subgroups (north – south) will provide us useful conclusions about the different effects the constraints have had on rich north and poor south. Lastly, we will evaluate whether the national fiscal rules can counteract political indiscipline and their efficiency on budget outcomes.

In the second section, we will approach the recent reform of the European governance, providing details of the reform package and discussing its appropriateness and its potential effectiveness on the restore of the European financial stability.

2. Discretionary Fiscal Policy in EMU – An Empirical Approach

The main objective of the empirical analysis is to examine to what extent the constraints of both Maastricht criteria and Stability and Growth Pact (SGP) have affected the way national authorities conduct their discretionary fiscal policy. Provided that the monetary policy of all countries in the Eurozone is managed by the ECB, the fiscal policy undertakes the responsibility to operate as a stabilizing tool of the business cycle and to counteract the negative asymmetric shocks. Consequently, it is the foremost tool in the quiver of governments to deal with their country-specific fluctuations. For this reason, we would expect that the process of European integration should be linked to the adoption by the member-states of more countercyclical discretionary fiscal policies. On the other hand, the existence of the Pact sets constraints and limitations on the conduct of fiscal policy. The question we want to answer is whether these constraints prevent the stabilizing role of fiscal policy and if this hypothesis is supported by the empirical findings.

Making clear what the stabilizing role of authorities means, the governments tend to implement restrictive monetary and fiscal policies during booms and loose policies during recessions so as to stabilize the cycle. A rational assumption is that we should expect that European monetary union would be associated with the conduct of more strongly countercyclical fiscal policies which will affect negatively the budget outcome in times of economic recession as this is the way fiscal policy plays a stabilizing role in business cycles.

This analysis is based on that of Gali and Perotti (2003) aiming to amend and extend it. Specifically, we use historical data until the year of 2010 and we add in our model the variable of national fiscal rules. The latter enables us to evaluate whether the national fiscal rules can counteract political indiscipline and provide balanced budget outcomes. The division of the EZ countries into two subgroups (north – south) will provide us useful conclusions about the different effects the constraints have had on rich north and poor south.

From a methodological perspective, our empirical approach focuses on the variables that constitute indicators of discretionary fiscal policy such as the structural deficit or cyclically unadjusted deficit. It is essential to make a distinction between the changes in fiscal policy that occur as specific measures decided by national authorities discretionarity and the changes as a result of the general economic conditions that affect the automatic stabilizers. The level of the deficit consists of the cyclical deficit which is the result of business cycle fluctuations rising during recessions and falling during booms since the cyclical deficit acts as an automatic stabilizer and the structural deficit which shows how large the deficit would be if the economy were operating at full employment (potential real output) and demonstrates the impacts of the actions adopted by the national authorities whose objective is to cope with the endogenous or exogenous (such as the financing of a war) fluctuations of the cycle. A typical example of the cyclical deficit is the reduction of tax revenues and the increase of payments for social insurance during recessions.

2.1 Methodological Framework

The first step is to examine the stationarity characteristics of each time series. Actually, there are numerous econometric techniques to test for the existence of a unit root. In the current study, we use the popular Augmented Dickey – Fuller methodology (ADF) (Dickey and Fuller, 1979).¹

The ADF test is based on the following regression (Kaskarelis 1993):

$$\Delta Y_{t} = a + bt + \rho Y_{t-1} + \sum_{i=1}^{m} \gamma \Delta Y_{t-i} + \varepsilon_{t}$$

where Δ is the first difference operator, t is time and ε_t is the error term.

In case the cyclical component is stationary, the secular component has a unit root and Y follows a random walk process i.e. the change in Y is absolutely random. Algebraically a random walk has the following form: $\mathbf{Y}_t \cdot \mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t$. Furthermore, if $\alpha \neq 0$, then Y follows a random walk process with a drift. A drift process is represented as follows: $\mathbf{Y}_t = \mathbf{Y}_{t-1} + \alpha + \boldsymbol{\varepsilon}_t$. Note that the lag dependent polynomial is incorporated with the aim to deal with the potential serial correlation of the residuals.

However, it is well-known that regarding panel data series, the standard unit root tests based on individual time series are not the appropriate techniques to employ as they do not work effectively. This is why we tend to apply panel data unit root tests that are employed in the investigation of statistical properties in panel data analysis. The results provided by the panel data unit root tests will be more reliable since the panel data analysis increases the strength of the test by enhancing the time series dimension of the data by the cross section. There are several panel unit root tests, some of the most popular are the following: the ADF - Fisher Chi-square (Maddala and Wu,1999), PP – Fisher Chi-square (Choi, 2001), the LLC (Levin, Lin and Chu, 2002)

¹ There are several unit root tests that can be used such as the test of Zivot and Andrews (1992), the IPS test (Im et *al.* 1997), the MW test (Maddala and Wu, 1999), or the Choi test (Choi, 2001).

and the IPS (Im, Pesaran and Shin, 2003)². For our analysis, we use the method of ADF – Fisher Chi-square as an alternative approach to the unit root tests. The ADF – Fisher Chi-square test combines the p-values from the individual unit root tests and allows for individual unit root processes so that p-values vary across cross-sections.

The ADF - Fisher Chi-square is based on the following regression (Baltagi, 2001; Fischer, 1932):

$$P = -2 \sum_{i=1}^{n} lnp_i$$

The hypothesis that we have to evaluate is H_0 : $\rho i = 1$ against the alternative H_1 : $\rho_t < 1$ (the series are weakly stationary or trend stationary). The ADF - Fisher Chi-square test was applied both on the initial original variables of the models and their first differences. Most of the original variables are non-stationary however their first differences are stationary.

Moreover, in order to choose the appropriate coefficient covariance method, we work in full accordance with the Arellano asymptotics (1987). If T (number of periods) is greater than N (number of cross sections) and T<2N we use the method of White diagonal with Cross Section weights, while if T>2N we use the method of White Cross section with Cross Section SUR weights. As a result, for models 2,4 we use the method of White diagonal while for models 1,3, the method of White Cross section.

Table 2 shows the results of the ADF-Fischer Chi-square test for panel data. The ADF test was applied both on the original variables and their first differences. Most macroeconomic variables are non-stationary; however all their first differences are

² While the LLC test allows for heterogeneity of individual deterministic effects and a heterogeneous serial correlation structure, it assumes the presence of a homogeneous autoregressive root under the alternative. The latter is identified as a serious limitation for the LLC test. The LLC test procedure involves using pooled t-statistics of the estimator to evaluate the hypothesis of non-stationarity of each individual time series. The more recently developed IPS tests overcame the limitation of the LLC test by allowing for heterogeneity of the autoregressive root under the alternative. The IPS test is simple to calculate and allows for residual serial correlation and heterogeneity of dynamics across groups. However, simulations indicate that the IPS test is sensitive to a correct choice of lag orders in the underlying ADF regressions; the power of the t-bar test is more favorably affected by a rise in time dimension of the data than the cross-section units of the data; and the interpretation of the IPS test results are difficult because of the heterogeneous nature of the alternative hypothesis. Maddala and Wu's (1999) and Choi's (2001) tests were similar in the way that both suggested panel unit root tests performed using a Fisher statistic, but they were developed to overcome the shortcomings of the LLC and the IPS tests. Maddala and Wu's (1999) and Choi's (2001) tests solves the problems related to previously mentioned tests by providing the combination of probability values for a unit root tests applied to each group in the data set. With this in mind, we employed the LLC, the IPS, ADF-Fisher and PP-Fisher panel unit root tests in this paper. For the LLC and IPS test, the optimal lag length is determined according to Schwarz criteria.

stationary. Finally, the table 3 shows the cross section effects for our main specification.

Table 2: ADF-Fischer Chi-square statistics

Deficit Original Variables

Method	Statistic	Prob.	sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t	-1.61878	0.0527	10	258
Im, Pesaran and Shin W-stat	-2.39357	0.0083	10	258
ADF - Fisher Chi-square	34.6693	0.0219	10	258
PP - Fisher Chi-square	30.1216	0.0679	10	268

Output Gap Original Variables

Method	Statistic	Prob.	sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t	-2.95752	0.0016	7	182
Im, Pesaran and Shin W-stat	-5.58379	0	7	182
ADF - Fisher Chi-square	58.5476	0	7	182
PP - Fisher Chi-square	25.7609	0.0278	7	189

Adjusted Deficit 1st differences

Method	Statistic	Prob.	sections	Obs
Null: Unit root (assumes comme	on unit root pro	cess)		
Levin, Lin & Chu t	-5.55320	0	10	242
Im, Pesaran and Shin W-stat	-7.07650	0	10	242
ADF - Fisher Chi-square	89.5749	0	10	242
PP - Fisher Chi-square	161.657	0	10	252

Lagged Output Gap 1st differences

Method	Statistic	Prob.	sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t	1.94074	0.9739	6	152
Null: Unit root (assumes individual unit root process)				

Im, Pesaran and Shin W-stat	-3.07318	0.0011	6	152
ADF - Fisher Chi-square	29.5077	0.0033	6	152
PP - Fisher Chi-square	15.4438	0.2181	6	158

Fiscal Rule Index 1st differences

Method	Statistic	Prob.	sections	Obs
Null: Unit root (assumes commo				
Levin, Lin & Chu t	-6.03639	0	5	85
Im, Pesaran and Shin W-stat	-4.03450	0	5	85
ADF - Fisher Chi-square	35.6051	0.0001	5	85
PP - Fisher Chi-square	44.7099	0	5	90

Lagged Gross Debt 1st differences

Method	Statistic	Prob.	sections	Obs
Null: Unit root (assumes commo				
Levin, Lin & Chu t	0.91956	0.8211	7	171
Im, Pesaran and Shin W-stat	-2.31178	0.0104	7	171
ADF - Fisher Chi-square	27.8626	0.0148	7	171
PP - Fisher Chi-square	31.1387	0.0053	7	178

Lagged Adjusted Deficit 1st differences

Method	Statistic	Prob.	sections	Obs
Null: Unit root (assumes common uni				
Levin, Lin & Chu t	-1.98181	0.0238	5	125
Im, Pesaran and Shin W-stat	-5.84108	0	5	125
ADF - Fisher Chi-square	52.9095	0	5	125
PP - Fisher Chi-square	113.115	0	5	130

Austria	0.237652
Belgium	-0.769340
Finland	1.213371
France	0.221274
Germany	0.331056
Greece	-0.686844
Ireland	-0.597951
Italy	-0.797347
Netherlands	0.159422
Portugal	0.428694
Spain	0.331056

Table 3: Cross Section Effects (Equation 3)

2.2 Panel Data Regressions and Empirical Analysis Results

Our sample consists of the data of the 11 initial members – states of the Eurozone (Austria, Belgium, Finland, France, Germany, Greece³, Netherlands, Ireland, Italy, Portugal, and Spain). The data are on annual basis, come from the database of OECD and cover the time period 1981-2010, and capturing traces of the current crisis.

A useful starting point for our empirical analysis would be to regress the following relation:

$$\mathbf{d}_{\mathbf{t}} = \mathbf{a}_{\mathbf{0}} + \boldsymbol{\beta}_{\mathbf{x}} \mathbf{X}_{\mathbf{t}} + \mathbf{d}_{\mathbf{t-1}} + \mathbf{u}_{\mathbf{t}} (1)$$

where \mathbf{d}_t is the deficit of general government as a share of GDP, \mathbf{X}_t is the output gap and \mathbf{d}_{t-1} is the lagged variable of deficit.

The concept is to regress an indicator of fiscal policy on a cyclical indicator, so we will estimate the relation between the cyclically unadjusted deficit of general government and the output gap which is an economic measure of the difference between the actual output of an economy and the potential output (the output that can be produced at full employment). The use of the lagged variable helps us to account for the likely of error autocorrelation and it allows explanatory variables to have effects beyond the current period.

³ Greece joined EMU in 2001.

Even though this relation does not identify the systemic response of national authorities as discretionary policy to the fluctuations of the cycle, it provides a useful descriptive relation between public finances and cyclical activity. Our results demonstrate the contribution of cyclical conditions on the implementation of balanced or surplus budgets and hence on the ensuring of the sustainability of public debt.

The table 1 displays the results for our specification. Even if our model is simplistic, it has an appealing interpretative capacity. The explanatory variables are statistically significant at the significance level of 95%. Particularly, the results demonstrate a clear positive relation between the level of cyclically unadjusted deficit and the output gap. A reduction in the negative output gap or an increase in the positive output gap by 1%, would reduce the level of deficit by 0,5%. It would be wrong to conclude that the national authorities tend to conduct procyclical fiscal policy due to the fact that we have not used the appropriate indicators of discretionary policy in our specification.

Interpreting the empirical results of the model 1, they highlight the weaknesses in the structure of SGP. Regarding the Excessive Deficit Procedure (EDP) which entangles the imposition of fines in case there is a deficit in excess of 3% of GDP, we approve of the reviews which state that the SGP restricts the necessary flexibility fiscal policy should have in order to stabilize the cycle. Moreover, the SGP should take into account the growth rate of member-states and also their position into the business cycle since the rule refers to the cyclically unadjusted deficit (debt dynamics equation: $g - t + (r - x)b = \mathbf{b}$) and to the structural deficit.

In order to examine how authorities utilize fiscal policy as a tool to stabilize the fluctuation of business cycle, we use the structural deficit as an indicator of fiscal policy stance. Firstly, we should determine properly the timing of fiscal policy decisions so as to define the nature of the variable the national authorities react to. Actually, the measures are usually decided approximately a year before their implementation, excluding exceptional cases. Therefore, national authorities' decisions should be based on the expectation of the output gap, conditional available on information available in the period t-1 ($E_{t-1} X_t$). However, reality proves that the process of policy making is characterized by complexity and inconsistency, so a plausible assumption would be that the structural deficit responds to the output gap in the period t-1, rejecting a forward looking approach. Furthermore, in our model we incorporate the variable of the measure of gross debt relative to potential output gap as a debt stabilization motive (Gali and Perotti, 2003; Bohn, 1998; Wyplosz, 2002) and the variable of the lagged dependent variable (by one year) in order to avoid autocorrelation error and to deal with endogeneity possibilities⁴. The introduction of

⁴ Dealing with the problem of endogeneity is a complicated task. In econometric theory, it is vague whether a variable is endogenous or exogenous. It depends on the assumptions made by

these two explanatory variables enables us also to take into account the initial limitations faced by the government. The resulting specification we estimate is the following:

$$\mathbf{d}_{\mathbf{t}}^{*} = \mathbf{a}_{\mathbf{0}} + \beta_{\mathbf{x}} \mathbf{X}_{\mathbf{t}-1} + \beta_{\mathbf{b}} \mathbf{b}_{\mathbf{t}-1} + \mathbf{b}_{\mathbf{s}} \mathbf{d}_{\mathbf{t}-1}^{*} + \mathbf{u}_{\mathbf{t}} (2)$$

where \mathbf{d}_{t}^{*} is the structural deficit divided by potential output, \mathbf{X}_{t-1} is the output gap for the period t-1, \mathbf{b}_{t-1} is the gross debt of general government as a share of GDP for the period t-1 and \mathbf{d}_{t-1}^{*} is the lagged dependent variable.

A negative (positive) value of the coefficient β_x implies that fiscal authorities use discretionary fiscal policy in a countercyclical (procyclical) way. A negative value of the coefficient β_b , as well as a value of the coefficient \mathbf{b}_s less than 1, implies that policymakers are subject to initial restrictions regarding the level of deficit and debt (Gali and Perotti, 2003). The higher the initial level of debt or deficit, the lower they conduct strongly countercyclical discretionary policy. Since our primary objective is to detect whether the constraints of Maastricht criteria and SGP have impaired the way policymakers conduct discretionary fiscal policy, we split our sample into two sub periods: the pre-Maastricht period and the post-Maastricht period. The first sub period covers observations for the period from 1981 to 1991 (one year before the criteria of Maastricht Treaty come into force). The empirical results for this period will demonstrate the tendency of policymakers in fiscal policy making process and how they conduct discretionary policy without constraints and limitations. We estimate the following version:

$$\mathbf{d}_{t}^{*} = \mathbf{a}_{0} + \boldsymbol{\beta}_{xBM} \mathbf{X}_{t-1} + \boldsymbol{\beta}_{b} \mathbf{b}_{t-1} + \mathbf{b}_{s} \mathbf{d}_{t-1}^{*} + \mathbf{u}_{t} (2a)$$

where the initials BM and AM refer to pre-Maastricht and post-Maastricht periods respectively.

Looking at the results of the model 2a from the table 1, in the pre-Maastricht period when governments had at their disposal also the monetary policy as a stabilizing tool, they tended to utilize the tools of fiscal policy in a systemic countercyclical way. The coefficient of output gap has a negative value which indicates that policymakers conduct restrictive fiscal policy during booms and loose fiscal policy during recessions. As far as the initial restrictions are concerned, we notice that initial limitations exist only in respect of the initial level of deficit, while the higher the initial debt, the lower the structural deficit national authorities set discretionarily. The

the analyst and his theoretical background. A way to deal with the "fear" of endogeneity is to use an instrumental variable which allows consistent estimation when the dependent variable causes at least one of the explanatory variables. That means that there is a reverse causation and our results are biased.

magnitude of the gross debt does not constitute a deterrent factor for the adoption of countercyclical fiscal policy. Note that both the model and the independent variables are statistically significant at the 0.05 level.

The second sub period under examination covers the period from 1992 (when the criteria of Maastricht came into force regarding the membership in the Eurozone) to 2010 including the effects of the adoption of the supranational rule for EZ member states.

$$\mathbf{d}_{t}^{*} = \mathbf{a}_{0} + \mathbf{\beta}_{xAM} \mathbf{X}_{t-1} + \mathbf{\beta}_{b} \mathbf{b}_{t-1} + \mathbf{b}_{s} \mathbf{d}_{t-1}^{*} + \mathbf{u}_{t} (2b)$$

The results of the analysis support our hypothesis that the integration of monetary policy with a clear mandate to the focus on the target of price stability is associated with countercyclical fiscal policies in the EMU countries even if the flexibility of fiscal policy is being reduced when the medium-term target of the SGP has not been achieved. Nevertheless, even though the explanatory variable is not statistically significant at level lower than 20%, there is an indicative tendency of a significant reduction in the degree of counter-cyclicality of discretionary fiscal policy. Additionally, it is concluded that the supranational fiscal rule for the level of deficit has significantly limited the capacity of policymakers to use fiscal policy a stabilizing tool of the cycle as the empirical data confirm the failure of member-states (especially France and Germany) to comply with the rule.

Now, we repeat the same exercise, having divided our sample into two sub groups. We split our sample of countries into the poor south or PIGS (including Greece, Italy, Spain, Portugal and Ireland) and the rich north (including Germany, France, Finland, Austria, Belgium and Netherlands). This will enable us to extract the different features and the asymmetries between the two sub groups as far as the conduct of fiscal policy is related. The pattern that emerges, shows that the southern European countries run systematically countercyclical discretionary fiscal policies in the post-Maastricht period which is statistically significant at 0,05 level, but there is a reduction in the degree of counter-cyclicality from the pre-Maastricht period which is statistically significant at 0,10 level. On the other hand, regarding the northern countries, they appear to conduct procyclical discretionary policies in the post-Maastricht period in contrast to the previous when there is a statistically significant negative relation between structural deficit and output gap. The above finding demonstrates an aspect of the decreasing synchronization among the counterparts of the Eurozone⁵.

⁵ Papageorgiou et al. (2010) testify a decreasing synchronization among the counterparts of the emu zone after the introduction of the euro coin".

Following the lead of several authors, we also incorporate into our model the independent variable of national numerical fiscal rules (Iara and Wolff, 2010; Debrun et al., 2008 Ayuso-i-Casals et al., 2006; Commission, 2007; Deroose et al., 2006). Apart from the rules imposed by the SGP, there are numerous national fiscal rules which are designed to prevent the decline of public finances and to hit the profligacy of governments. A concise definition of the national fiscal rule is the one proposed by Kopits and Symansky (1998) which defines the national fiscal rule as "a permanent constraint on fiscal policy, expressed in terms of a summary indicator of fiscal performance". In order to meet the needs of the scientific research, Commission firstly compiled a dataset on national fiscal rules in force across EU countries and then created the Fiscal Rule Strength Index which evaluates numerically the strength and the efficiency of domestic fiscal rules. Five criteria have been taken into consideration: the statutory/legal base of the rule, the room for setting or revising objectives, the nature of the body in charge of monitoring respect and enforcement of the rule, the enforcement mechanisms of the rule and the media visibility of the rule⁶. The ranking of the index takes values from -1,12 to $1,54^7$. The use of the variable of national fiscal rules enables us to evaluate the contribution of domestic restrictions on the conduct of balanced budgetary outcomes and to what extend their strength affects the level of the structural deficit produced. The resulting specification that we estimate is thus:

$\mathbf{d}_{t}^{*} = \mathbf{a}_{0} + \boldsymbol{\beta}_{x} \mathbf{X}_{t-1} + \boldsymbol{\beta}_{b} \mathbf{b}_{t-1} + \mathbf{b}_{s} \mathbf{d}_{t-1}^{*} + \boldsymbol{\beta}_{f} \mathbf{f}_{t} + \mathbf{u}_{t} (3)$

The most natural interpretation of the above findings is that there is an undeniably positive relation between domestic fiscal rules index and the level of structural deficit. The higher the fiscal rule strength index is for a country, the greater contribution of domestic constraints on the level of deficit produced. However, this relation is not statistically significant at a level lower than 25%. Moreover we find that the presence of national numerical fiscal rules increase the extent of countercyclicality of fiscal policy. Finally, it must be noted that there is a strong negative relation between the output gap for the period t-1 and the structural deficit for the period which proves one more time that governments run strongly countercyclical discretionary fiscal policy.

⁶http://ec.europa.eu/economy_finance/db_indicators/fiscal_governance/documents/fiscal_rules _calculation_fiscal_rule_index_2010.pdf

⁷http://ec.europa.eu/economy_finance/db_indicators/fiscal_governance/fiscal_rules/index_en.ht m

Independen t variables	Model 1	Model 2 (a-b)		Model 2 (southern countries)	
Output gap t-1		-0.167585 (-2.017307)*	-0.047525 (-1.243943)	-0.255622 (-1.823910)**	-0,089698 (-1,917188)**
Gross debt t-1		0.087848 (2.383029)*	0.024174 (3.791262)*	0,151052 (5,757249)*	0,027754 (3,016101)*
Lagged adjusted deficit		0.288348 (2.084699)*	0.813873 (19.52911)*	0,336790 (2,133473)*	0,812107 (15,27963)*
Output gap	0.499090 (5.510424)*				
Deficit t-1	0.715293 (9.988716)*				
Fiscal Rules Index					
Constant	-0.741778 (-3.126519)*	-9.828870 (-3.392398)*	-2.192489 (-4.494069)*	-17,11492 (-8,660905)*	-2,889977 (-3,542233)*
R2	0.759422	0.841258	0.874724	0.694813	0,838122
Durbin- Watson stat	1.530709	1.612692	1.856546	1.559896	1,693448
F-stat	75.08105	32.23887	104.7353	8,130985	64,34871
Countries included	11	10	11	5	5
Total panel observation s	283	86	209	33	95
Coef. Covariance Method	White Cross section	White diagonal	White diagonal	White Cross section	White Cross section
Period	1981 - 2010	1981 - 1991	1992 - 2010	1981 – 1991	1992 - 2010

Table 1: Panel Data Regression Results

Independent variables	Model 2 (northern countries)		Model 3
Output gap t-1	-0.131373	0.065972	-0.092046
	(-2.161651)*	(1.050242)	(-2.108149)*
Gross debt t-1	0.012280	0.015535	0.019723
	(0.639387)	(1.723184)**	(2.666155)*
Lagged adjusted	0.363182	0.7443	0.823360
deficit	(4.012788)*	(11.70722)*	(20.25733)*
Output gap			
Deficit t-1			
Fiscal Rules Index			0.202563
			(1.152632)
Constant	-3.303735	-1.329439	-1.823903

	(-2.4966613)*	(-2.037808)*	(-3.387655)*
R2	0.852416	0.839483	0.872487
Durbin-Watson stat	2.263396	2.062901	1.928381
F-stat	37.13004	68.64201	91.88301
Countries included	5	6	11
Total panel observations	53	114	114
Coef. Covariance Method	White Cross section	White Cross section	White diagonal
Period	1981 - 1991	1992 - 2010	1992 - 2010

In parenthesis are depicted the t-stat values. Model 1: dependent variable is the cyclically unadjusted deficit as a share of GDP. Model 2,3: dependent variable is the structural deficit as a share of potential GDP. * the independent variable is statistically significant at 0.05, ** the independent variable is statistically significant at 0.10.

3. The Reform of European Governance framework – A Critical Review

The global financial crisis revealed not only the member states' real fiscal state but also the undercover weaknesses of the European institutional framework of European governance. In 2010, the global financial crisis reverted into a debt crisis for the entire euro area which is now characterized systemic. In order to overcome the impacts of the crisis and to restore stability, the leadership of the EU decided to adopt radical measures so as to shield the European governance is considered as a decisive step towards the fiscal and political integration. The paper asks whether these reforms seem sufficient and appropriate to deal with the matter of the restore of stability in the EMU.

Concisely, the reform package contains the following:

- A stronger version of the Stability and Growth Pact associated with more severe enforcement mechanisms and sanctions, and also the introduction of a new debt criterion under the Excessive Deficit Procedure (corrective arm).
- A new directive on national budgetary frameworks which entangles the setting of minimum requirements in consistency with fiscal framework of the EU into the domestic legislation and the adoption of the golden rule for balanced national budgets (over the cycle).
- A robust framework for preventing and correcting macroeconomic imbalances. The introduction of the Excessive Imbalance Procedure aims to identify and alleviate the macro-economic imbalances via the use of an indicative scoreboard, combined with techno-economic judgment.

- The enhancement of economic coordination with several interlinked and coherent policies for sustainable growth, convergence and high-level competitiveness through the implementation of the Euro Plus Pact.
- The establishment of a permanent European Stability Mechanism (ESM) which aims to offer financial assistance to member-states incapable of accessing market financing. Access to ESM will be provided on the basis of strict economic policy conditionality, under an adjustment programme and a rigorous analysis of public-debt sustainability. ESM will coexist with the EFSF and their overall financial capacity will be 500 billion Euros.

A rational assumption is that the efficiency of the reform package depends on the addressing of the factors that caused the crisis. What we can conclude from the nature of the reform is that the major problem which must be addressed is the profligacy of national authorities and the fiscal indiscipline. This is why the leadership of the EU promoted inter alia the enhancement of the SGP with both stricter rules and enforcement mechanisms. Upon this, a question arises: *is the fiscal indiscipline the root of the European sovereign debt crisis*?

3.1. Wrong approach

Data analysis shows that the period before the outbreak of the global financial recession, the public finances for the euro area were highly satisfactory. Particularly, in 2007, the average deficit of the union reached 0,7%, demonstrating not only a great reduction of the level of deficit since the beginning of EMU but at the same time it was a record low since the 70s. Similarly, the average gross public debt declined from 73% in 1998 to 66,5% in 2007, while the level of public expenditures decreased by 9 percentage points (Uxo and Paul, 2011). Consequently, one conclusion that emerges is that the sovereign debt crisis is not the result of the inability of governments to run sound fiscal policies. The countries that have been afflicted more by the impacts of the debt crisis used to run prudent fiscal policies in compliance with the rules of the SGP (see Spain and Ireland). The great exception to this story is the case of Greece, whose governments manipulated the data, producing the "famous" Greek statistics. The main reasoning behind the deterioration of the public finances can be found in two factors. Firstly, in order to alleviate the consequences of the global financial crisis, governments took the responsibility to save the banking system providing both liquidity and guarantees and at the same time they were forced to sustain economic activity at pre-crisis levels through the conduct of loose discretionary fiscal policy (De Grauwe, 2010). These actions undertaken by the national authorities, illustrate both the inherent flaws of the financial sector and the failure of monetary policy which did not manage to apply stricter controls on the functioning of banking system or to provide a regulatory framework so as to ensure stability. The recession appears thus to be the result of the unsustainable private debt explosion which forced governments to protect the private sector from the impacts of bubbles the financial sector itself caused. As a

result, what we can conclude is that the sovereign debt crisis has been created by a combination of complex factors and that the deterioration of public finances seems to appear more as a consequence of the crisis and not its fundamental cause.

3.2. Stricter SGP

The above conclusion raises doubts over the appropriateness of a more rigorous rule (as the one Commission introduced) whose objective is to counteract the current sovereign debt crisis. But beyond this, there are also problems which have been identified regarding the expected efficiency of the SGP and its structure in general.

To begin with, in the short term, the direct implementation of the new provisions will bring about a wave of measures of austerity which contain huge social cost and will result in the deepening of the recession with dramatic consequences for the real economy. This is why the adoption of radical measures will undeniably reduce the aggregate demand and will lead to a prolonged decline while there will not be any available tool to restore the economy. This is what is called vicious cycle. Moreover, the reduction in government expenditure will lead to a great fall in activity that gross debt ratio will get even worse, at least in the short term. This assumption is based on the fact that the cutting on spending will lower the GDP and as a result the gross debt ratio will increase. This will occur only if the multiplier (Keynesian models) is higher than 1.0 which means that if the government reduces its deficit by 1% of GDP by cutting expenditures while the multiplier is 1.8, the spending cut will lower GDP by 1,8% and the debt ratio will augment (Gros, 2011).

The tightening of the rules and the automaticity of enforcement mechanisms are going to reduce the flexibility that discretionary fiscal policy should have. Full compliance with the mid-term fiscal consolidation plans and the assessment on the basis of expenditure developments will definitely disable the capacity of fiscal policy to be used as a stabilizing tool for the fluctuations of the business cycle. On the other hand, the fact that building on the SGP, budget finances must be expressed in structural terms which seems to be in the right direction as it may be proven a mean capable of coping with the reasons which provoke the tendency for deficit bias⁸. To sum up, the new revised SGP predicts the conduct of countercyclical discretionary fiscal policy only in good times and leaves little room for countercyclicality during negative shocks. The necessary balance between flexibility and discipline is not achieved.

Structural problems are not avoided. The new obligation for member-states to keep debt below or sufficiently declining towards 60% of GDP seems excessively ambitious

⁸ Calmfors and Wren-Lewis (2011) distinguish six reasons for deficit bias: (i) impatience, (ii) informational problems, (iii) common pool problem, (iv) electoral competition, (v) exploiting future generations and (vi) time-inconsistency problem.

especially for countries subject to high gross public debt. Tamborini's technical analysis (2011) demonstrates the degree of difficulty of the above target. Tamborini attempts to evaluate the new SGP rules by means of dynamic models of the debt/GDP ratio focusing on its determinants: the real growth rate, the inflation rate and the nominal interest rate on the accumulated debt stock. He finds out that the convergence of debt/GDP ratio at 60% and the target of keeping it stable over time can be achieved if only if the prerequisite of uniform growth, interests and inflation rates for the entire euro area is satisfied. This appears quite difficult since the interest rate convergence seems impossible to be achieved due to the historical differences in risk measures. Finally, he concludes that heterogeneity and interdependence entails different features for m-s regarding the chances of success, the efforts and the spillovers towards the SGP target which finally seems infeasible. The appendix provides a useful approach of the nature of supranational rules concerning the objective of the debt reduction.

The revised SGP entangles the assignment of national responsibilities (decisions on tax policy and public spending) to EU institutions. The role of supranational institutions becomes more extensive and gains in power. In particular, the Commission may impose sanctions under the jurisdiction of the European Court of Justice, it may ask from member states to reduce public expenditure or to adjust their tax policy. This entails firstly that there is a democratic deficit in EU affairs and secondly that member-states should surrender their national sovereignty to the EU. These procedures also lack legitimacy as the European supranational institutions do not face the democratic political sanctions of their decisions and actions⁹ (De Grauwe, 2010).

It is worth mentioning that there are reviews which are in favor of the content of the revised SGP but they pose concerns for its efficiency. Their main concern is the lack of automatic sanctions. Even though the penalties will have a much greater degree of automaticity via the use of reverse QMV¹⁰, the lack of automatic sanctions will allow member-states to avoid their obligations under the Pact, increasing the room of dicretionary maneuvers for the national authorities in the European Council (Von Hagen, 2010). Fuest (2011) states that the reform package gives a lot of emphasis on the coordination and supervision approach instead of introducing more compliance via automatic sanctions and enforcement mechanisms. Schuknechat et al. (2011) asks for greater independence for the Commission services in its administration of the Pact and restrictions in member states right for veto.

⁹ There cannot be taxation without representation.

¹⁰ According to the so-called reverse qualified majority vote, the decision on the sanctions will be adopted unless a qualified majority of the Council vote against it.

3.3. Competitiveness and the Excessive Imbalance Procedure

It is generally admitted that fiscal policy does not guarantee by itself the fiscal stability and therefore should not be examined in isolation. There should be a broader macroeconomic surveillance to ensure the sustainability of public finances. The existence of excessive macroeconomic imbalances including divergences in current account and competitiveness directly affects the public finances and makes memberstates public finances more vulnerable to negative shocks. In order to shield memberstates from macroeconomic imbalances, the EU leadership decided to introduce a completely new procedure for excessive imbalance, the Excessive Imbalance Procedure. The philosophy underpinning the new provision seems to be the correction of trade imbalances across the EU through the enhancement of competitiveness.

However, the direction of the approach adopted seems to have serious weaknesses which may jeopardize the functioning of single market. The correction of the current account imbalances does not require only the increase of the competitiveness of countries with high-level current account deficits but also that deviations in both directions should be addressed since the side of surplus countries have a great margin of fiscal maneuvers. A symmetric approach seems more appropriate to cope with the solution of imbalances within the EMU (Uxo and Paul, 2011; Belke, Schnabl and Zemanek, 2010; Goodhart and Tsomocos, 2010; Pisany-Ferry, 2010; Dadush and Eidelman, 2010; Stockhammer, 2010). On the contrary, the Commission decided to promote an asymmetric approach which is based on the reasoning that current account deficits or trade deficits are the results of the lack of competitiveness, and surpluses come from the high-level economic efficiency, hence aiming at the improvement of competitiveness for the entire euro area (Wyplosz, 2010; Budesbank, 2010). Notwithstanding, the asymmetric approach ignores the fact that there is an intense interdependence among the member states as far as the trade is related. The structure of EU trade (intra-community trade mainly) suggests that if a member-state increases its competitiveness, another country must lose competitiveness. Overall improvement of EU competitiveness can be achieved only if the terms of trade with the rest of world become better.

Furthermore, it is generally accepted that macroeconomic imbalances within a monetary union are not necessarily negative as they demonstrate an improvement of the regional allocation of capital (Von Hagen, 2010). Also, changes in competitiveness reflect the convergence of standards of living for certain countries in the euro area. Finally, it is likely that under the procedure of trade imbalances correction there may be conflicts with other areas of EU policy such as regional development policy, labor market policy, single market policy etc.

Finally, technical problems are unavoidable. The competitiveness of an entire economy is very difficult to determine as the relative prices and the remunerations depend on many aspects of labor and capital market institutions. Evaluating and enhancing competitiveness constitutes a difficult task to deal with since many criteria which define the production cost must be examined, not only the aspect of direct labor cost on which all attention has focused. Lastly, the determination of specific thresholds for each member-state beyond which an existing imbalance becomes crucial and dangerous lacks technical methodology and seems to be unreliable and abstract (Belke, 2010).

3.4. Need for financial sector reforms

Taking into account that the major causes of the sovereign debt crisis in the Eurozone are the accumulation of private debt and the fiscal indiscipline, the reform of European governance should not only focus on the fiscal surveillance framework but also on the financial sector and on the conduct of monetary policy.

The adoption of a permanent mechanism that will help preserve the economic and financial stability of the Union itself by providing financial assistance and containing specific provisions for debt restructuring should be associated with fundamental reforms in the financial sector since debt and banking crisis are the two sides of the same coin. The reforms in the financial system should firstly provide clauses and guarantees in order to prevent the outbreak of a new banking crisis and secondly they should undertake corrective actions so that the financial sector may be sufficiently robust to absorb the impacts of a sovereign default or a potential debt restructure.

The accumulation of the private debt in banks is the greatest concern. European institutions should establish a regulatory framework that will govern banking transactions and monitor the operation of the banking system. In that framework, the sustainability of European banks should be assessed via austere stress tests¹¹ monitored and supervised by independent European institutions. Actually, a crucial parameter is the monitoring of economic indicators that may prevent the creation of bubbles into the financial sector. Authorities should identify the risks and send warning and alarms to national authorities so as they can get prepared for a potential collapse.

The Eurozone's permanent bailout fund with a lending capacity of 500 billion euro offers financial assistance only to member-states when their regular access to market financing is impaired, while the granting of loans and assistance to bank institutions depends on the discretionarity of the national authorities. From this perspective, if the ESM made room for the support from the banking system with extra capitals under specific terms, the moral hazard related with the rescue of banks by governments would be reduced significantly. Concerning the involvement of the private sector in

¹¹ Authorities have to deal with the problem of the undercapitalization of banks and how the latter ensure liquidity for the real economy.

debt restructuring, this should not be unique and exceptional regarding the Greek debt but standardized and identical procedures and provisions seem essential to be introduced. In addition to that, these provisions should be associated with guarantees and clauses which will indeed ensure the sustainability of the banking system. The main conclusion we may derive is that the reform package is judged insufficient to improve financial sector stability. This also applies for the Basel III process which contains specific measures for broader financial stability. We keep our doubts whether the measures provided will make the sector more stable and rigid¹² over the long run, thus raising economic growth. Furthermore, an ongoing issue of great concern is the debate related with the role of rating agencies and the operation of derivatives market¹³. In particular, investors can buy Credit Default Swaps (CDS) contracts referencing national debt without owing any national bond which implies that there may be financial speculative motives. As collective speculative actions may set the stability of the system in danger, EU should adopt measures to ban CDS speculation.

It is also considered essential that European authorities push towards a bilateral agreement with countries known as tax havens in order to get valuable fiscal information for the fight against tax evasion and corruption, giving European societies inter alia a sense of justice.

3.5. National Fiscal Rules

The global financial crisis unfolded the inherent weaknesses of public finances of many EZ countries leading to the outbreak of the ongoing systemic crisis for the entire euro area. The key determinant of the deterioration of public finances is that national fiscal authorities seem to be prone to a deficit bias, which brings about large and persistent deficit and accumulating public debts. The causes behind the phenomenon of deficit bias are rooted in political business cycle determinants and can be summarized as follows: impatience, informational problems, common pool problem, electoral competition, exploiting future generations and time-inconsistency problem (Calmfors and Wren-Lewis, 2011). Governments, having the primitive objective of their re-election tend to conduct procyclical discretionary fiscal policy instead of countercyclical which would be preferable (IMF, 2004) so as to maximise their possibilities to be re-elected. This results in overspending and accumulation of deficit and debt.

¹² There should be provisions which make sure that banks in countries undergoing a debt restructuring still have access to the liquidity of ECB or to refinancing through ECB (Fuest, 2011).

¹³ Derivative markets are investment markets for financial instruments that get their value, or at least part of their value, from the value of another security, which is called the underlier.

Literature suggests that the adoption of national fiscal rules can tackle the political indiscipline and the sources of deficit bias contributing inter alia to the implementation of sound fiscal policies. Ayuso-i-Casals et al. (2006) distinguish two types of fiscal rules: the numerical fiscal rules and the procedural rules. First, the numerical fiscal rules set targets or thresholds for budgetary aggregates. The aim in this case is to pose an ex ante permanent constraint on fiscal policy so as to promote credibility and discipline. Second, the procedural rules govern the procedure of the implementation of annual budget ensuring the transparency and the accountability. In general, the scope of fiscal rules is very broad. They usually target toward different directions including debt, expenditure, deficit, and sub-government goals but all of them comply with the primitive goal which is the promotion of fiscal sustainability. Several studies have noticed that the nature of fiscal rules has changed significantly over time. Countries prefer a combination of rules linked to sustainability instead of a single rule as it was in the past. The objective of curtailing the size of public sector and the use of structural target has also come to the fore (IMF, 2009). In recent years, data shows that more and more countries tend to rely on the effectiveness of fiscal rules to conduct fiscal policy and that the most common form of fiscal rules among OECD members are the debt, budget balance and revenue rules.

Empirical evidence demonstrates that the adoption of national fiscal rules is associated with sound fiscal policies and improved fiscal performance (Kumar et al., 2009; Debrun et al., 2008; European Commission 2006; Deroose et al., 2006 and Debrun and Kumar, 2007). The higher the share of governments' finances governed by numerical fiscal rules, the lower the level of public deficit produced. Fiscal rules lead ceteris paribus to balanced budgetary outcomes or lower deficits (Ayuso-i-Casals et al., 2006). In particular, empirical evidence shows that strict fiscal rules are linked to better primary budgetary outcomes. Moreover, expenditure rules manage to counteract the profligacy of governments since they internalize the cost of expenditures alleviating inter alia the common pool problem which is at the source of overspending. Experience indicates that fiscal rules are more effective as disciplinary mediums at the central level of government rather than in a decentralized framework since they appear to be highly correlated with fiscal stability. However, under certain circumstances, fiscal rules in sub-national level can provide a useful policy framework (Ter-Minassian, 2005). Furthermore, fiscal rules are considered to be a determinant for successful fiscal adjustment and consolidation (Guichard et al., 2007; Kumar et al., 2009). Empirical researches for EU countries show that stronger fiscal rules are associated with a greater likelihood for successful consolidation (European Commission, 2007). Also, it must be mentioned that the different features of the rules constitute key elements that determine their effectiveness on budgetary outcomes. Finally, Manasse (2006), Von Hagen et al (2001) and Gali and Perotti (2003) argue that the presence of numerical fiscal rules increase the degree of countercyclicality of discretionary fiscal policy.

4. Conclusions

This paper made an attempt to answer some crucial economic questions: a) whether the limitations of Maastricht criteria and SGP have impaired the ability of national authorities to run countercyclical discretionary fiscal policy in the EMU context approaching the time period 1981-2010 and b) whether the recent reform of the European governance framework seems sufficient to deal with the impacts of the current sovereign debt crisis and to restore stability.

Estimating the model adopted, several interesting conclusions emerge. Firstly, discretionary fiscal policy has become less countercyclical overtime as we have found a significant reduction in the degree of countercyclicality of discretionary fiscal policy. Secondly, there are differences in the manner the two sub groups of EZ countries conduct their discretionary fiscal policy. More precisely, the countries that form the PIGS are found to run to some extent countercyclical policies while the northern countries tend to conduct procyclical fiscal policies after the process of monetary integration. Finally, the results confirm the popular view that the adoption of national fiscal rules is associated with more sound fiscal policy and fiscal discipline. However, readers should take into account the limitations associated with the empirical analysis and not to overestimate the findings provided.

From our perspective, the recent reform does not confront the roots of the crisis. The adoption of a stricter SGP will deepen the decline and will reduce the flexibility of discretionary fiscal policy to be used as a stabilizing tool. The introduction of Excessive Imbalance Procedure seems to be in the right direction but a symmetric approach is needed. Lastly, we lay emphasis on the need for reforms in the financial sector in order the European economy to be shield from future crises.

Concluding, we want to stress that even though we have identified several weaknesses and flaws concerning the nature of the recent European governance reform and its appropriateness, we would rather to consider our remarks and the context of our critical review in general, as useful caveats to the debate opened about the future of EMU.

References

Arellano, M., (1987), Computing Robust Standard Errors for Within-Groups Estimators. Oxford Bulletin of Economics and Statistics, 49(4), 431–43.

Ayuso-i-Casals, J., Hernande, D., Moulin, L. and Turrini, A. (2006), Beyond the SGP - Features and Effects of EU National – Level Fiscal Rules. *European Comission*.

Baltagi, B.H. (2001), Econometric Analysis of Panel Data. 2^{nd} ed. Chichester , John Wiley and Sons, Ltd.

Bundesbank, (2010) On the problems of macroeconomic imbalances in the euro area. *MonthlyReport*, July, Deutsche Bundesbank.

Belke, A. (2010), Reinforcing EU Governance in Times of Crisis: The Commission Proposal and Beyond. *Diw Berlin*, Discussion Papers 1082.

Belke, A., Schabl, G, Zemanek, H (2010), Current account imbalances and structural adjustment in the euro area. *Journal of International Economics and Policy*, Vol. 7, 83-127.

Bundesbank, (2010), On the problems of macroeconomic imbalances in the euro area. *MonthlyReport*, July, Deutsche Bundesbank.

Calmfors, L., Wren-Lewis, S. (2011), What Should Fiscal Councils Do? *University of Oxford*, Discussion Paper Series No 537.

Bohn, H. (1998), The behaviour of US public debt and deficits. *Quarterly Journal of Economics*, August, 949-963.

Choi, I., (2001), Unit Root Tests for Panel Data. Journal of International Money and Finance, 20, p. 249–272.

Dadush, U., Eidelman, V., (2010), Germany: Europe's Pride or Europe's Problem? CarnegieMoscow Center

De Grauwe, P. (2010), Fighting the Wrong Enemy. Voxeu.org, May.

Debrun, X., Moulin L., Turrini, J. A.. Ayuso-i-Casals, Kumar M. S., (2008), Tied to the Mast? National Fiscal Rules in the European Union. Economic Policy, 299–362.

Deroose, S., Moulin, L. and Wierts, P. (2006), National Expenditure Rules and ExpenditureOutcomes: Empirical Evidence for EU Member States. *WirtschaftspolitischeBlaetter*, 1/2006, pp. 27–42.

Dickey, D. A., & Fuller, W. A. (1979), Distribution of the estimates for autoregressive time series with a unit root. Journal of the American Statistical Association, 74, 427–431.

European Commission (2007), Public Finances in EMU.

Fisher, R. A. (1932), Statistical methods for research workers. Edinburgh, Oliver and Boyd.

Fuest, C. (2011), Will the Reform of the Institutional Framework Restore Fiscal Stability in the Eurozone? *University of Oxford*.

Gali, J., Perotti, R. (2003), Fiscal Policy and Monetary Integration in Europe. *NBER*, Working Paper 9773.

Goodhart, C., Tsomocos, D., (2010), How to restore current account imbalances in a symmetric way. *Eurointelligence*.

Gros, D. (2011), Can austerity be self-defeating? Voxeu.org, November.

Hallerberg, M., Strauch, R., Von Hagen, J. (2007), The design of fiscal rules and forms of governance in European Union countries. *European Journal of Political Economy 23*, 338-359.

Iara, A., Wolff, B.G. (2010), Rules and risk in the euro area: Does rules-based fiscal rules governance contain sovereign bond spreads? Economic Paper 433.

Im, K. S., Pesaran, M. H., & Shin, Y. (1997), Testing for unit roots in heterogeneous panel. Department of Applied Econometrics, University of Cambridge.

Kaskarelis, I. (1993), Investigating the features of Greek business cycles. SPOUDAI, 43(1), 19–32.

Kopits, G., Symansky, S. (1998), Fiscal Policy Rules. IMF, Occasional Paper No 162.

Levin, A., Lin, C. F., and C. Chu (2002), Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties. Journal of Econometrics, vol. 108, pp. 1–24.

Maddala, G. S., & Wu, S. (1999), A comparative study of unit root tests with panel data and a new simple test. Oxford Bulletin of Economics and Statistics, 61, 631–652.

Papageorgiou, T., Michaelidis, P., Milios, G.J. (2010), Business Cycles Synchronization and Clustering in Europe (1960-2009). Journal of Economics and Business, vol. 62, issue 5, 419-470.

Perotti, R. (2011), The Austerity Myth: Gain Without Pain? *International Macroeconomics and Public Policy*, Discussion Paper Series No 8658.

Pisani-Ferry, J., (2002), Fiscal discipline and policy coordination in the Eurozone: Assessment and proposals. Paper for the European Commission President's Group for Economic Analysis.

Schuknecht, L., Moutot, P., Rother, P., Stark, J. (2011), The Stability and Growth Pact, Crisis and Reform. *ECB*, Occasional Paper Series No 129.

Stockhammer, E., (2010), Greek debt and German wages. The role wage policy and economic policy coordination in Europe. *Economic policy: in search of an alternative paradigm*, Middlesex University

Tamborini, R. (2011), The New Rules of the SGP. Threats from Heterogeneity and Interdependence. *University of Trento*.

Uxo, J., Paul, J. (2011), The reform of the economic governance of the eurozone: macroeconomic supervision and coordination. *University of Castilla*.

Von Hagen, J. (2010), The Sustainability of Public Finances and Fiscal Policy Coordination in the EMU. *CASE Network Studies & Analyses*.

Wyplosz, C., (2002), Fiscal discipline in EMU: Rules or institutions? mimeo, Graduate Institute for International Studies, Geneva.

Wyplosz, C., (2010), Germany, current accounts and competitiveness. *VoxEU*, 31th March.

Appendix

The appendix provides a useful descriptive interpretation of the supranational fiscal rules (of debt and deficit rules) as they are entangled in the Stability and Growth Pact from a macroeconomic perspective.

Note that the public deficit is defined as the rate of change of the level of accumulated gross national debt i.e. $\delta(t) = \dot{D}(t)$.

In case the total output of the economy Y(t) increases by a constant rate $g (= \overline{\Psi(t)})$

and the ratio of deficit as a share of GDP is constant, $\overline{Y(t)} = \alpha > 0$, then we can derive an expression that determines how the level of gross national debt as a share of GDP

changes $d(t) = (\overline{Y(t)})$. From the first order condition, we have:

$$\dot{d}(t) = \frac{\dot{D}(t)Y(t) - D(t)\dot{Y}(t)}{\Upsilon^{2}(t)} = \frac{\dot{D}(t)}{Y(t)} - \frac{D(t)}{Y(t)} \frac{\dot{Y}(t)}{Y(t)} = \frac{\delta(t)}{Y(t)} - d(t)g(t) =$$
$$= -gd(t) (1)$$

where a is the deficit as a share of GDP, d is the gross debt as a share of GDP and g is the constant rate of growth.

According to the supranational fiscal rules, for $a \le 0.03$ and $d(t) \le 0.6$, from the expression above we have the following results:

- \Rightarrow **d(t)** is higher than zero for g < 0.05
- \Rightarrow **d(t)** is zero for g = 0.05
- \Rightarrow *d***(t)** is lower than zero for g> 0,05.

In other words, in case the public deficit as a share of GDP is constant and lower than 3%, and the level of gross national debt as a share of GDP is lower than 60%, then the level of the debt will decline if only if the growth rate is higher than 5%. A numerical example can be described as follows: Greece (2007): 0,066 - (0,042*1,05) = 0,022 > 0, it implies that the national gross debt of Greece augmented.

Assuming that $\delta(t) = a Y(t) + r(t) D(t)$, where r(t) is the interest rate in the period t and under the conditions that r(t) is a positive function of gross national debt as a share of

GDP,
$$r(d(t) \text{ with } d \rightarrow -\infty$$
 $< g \text{ and } d \rightarrow \infty$ $> g$, we have:

$$d(t) = \frac{W(t) + V(t)}{Y(t)} - d(t)g =>$$

Ϋ́(t)

From our assumptions, It seems rational that there is some value $d=d^*$ for which we have $r(d^*) = g$. When $d^*(t) = d^*$, then r(d(t)) - g = 0 $d^*(t) = a$. Provided that the value of a is significantly low then there will be a value d^{**} whereby an equilibrium can be achieved for any initial value of $d^*(t) < d^*$. On the other hand, in case the value of a is significantly high then an equilibrium cannot be achieved and the gross national debt as a share of GDP would increase.

Fiscal decentralization in rural local governments in Mexico: Changes in Accountability and Entrepreneurship in the local government structures

Moreno, Flor

Dept. of Management, King's College London Email: flor_silvestre.moreno@kcl.ac.uk

Abstract. A major change in government is the shift of responsibilities from central to lower government levels. Autonomy is at the core of the argument as a driver of accountability and efficiency. The decentralization theory assumes that local actors will behave in a more accountable and entrepreneurial fashion. This research contributes to the understanding of the impact of decentralization policies in disadvantaged settings sharing light on the appropriate degree of autonomy that can avoid dangers. It answers the question whether fiscal decentralization in Mexico has been associated with changes in patterns of accountability and entrepreneurship, which tend to promote good governance in rural local governments. This will be addressed by analyzing financial, political, performance, and administrative variables constructed from country scale census, financial data and surveys from 1990 to 2008. In order to answer this question, concepts of accountability and entrepreneurship will be defined at the rural level based on qualitative techniques. The empirical evidence suggests that rural municipalities have been engaged on both accountability and entrepreneurial behavior. However, these changes seem to be encouraged neither by autonomy nor political elections but by type of financial resources decentralized.

Keywords: *decentralization, intergovernmental transfers, political election, accountability, entrepreneurship.*

1. Introduction

Decentralization is an international trend that reached its peak in the late 1980's and is, perhaps, 'the most important trend in governance around the world' (Freitag and Vatter 2008). This trend stems from the local government need of financing local public services. To do it so, revenues should be collected through different source of revenues such as taxes, borrowing and central government transfers. Such financial autonomy is expected to foster good governance at the local setting. This research is

aimed at analyzing the outcomes derived from the decentralization reform in the case of Mexico. It looks at changes in the political sceneries before and after the reform as well as internal changes within the local government structures. It argues that financial autonomy via intergovernmental transfers might lead to changes in accountability and entrepreneurship. The remainder of this paper is organised as follow. Section 2 focuses on the literature review sub-sectioned under the background of the Mexican decentralization reform and its political scenario. Section 3 describes the data and methodology divided into two subsections, before and after sceneries and changes in patters of accountability and entrepreneurship after the reform. Section 4 presents the results of the empirical analysis, and Section 5 concludes.

2. Literature Review

In Litvack, et al (1998, 2) the word decentralization means 'the transfer of authority and responsibility for public functions from the central government to subordinated or quasi-independent government organizations or the private sector'. The trend toward this policy appeared in the 1970's and by 2000, according to the World Bank, 95% of democratic countries as well as other non-democratic countries such as China have decentralized in some degree (Rodriguez-Pose and Sandall, 2008). The theory of decentralization is grounded on economic and political arguments. The economic arguments involve productive¹ and allocative² efficiency of public goods and services and the political arguments refer to the control mechanism over local agents in terms of productive and allocative efficiency (see Tiebout, 1956; Oates, 1972, 1999). Productive efficiency is envisaged as a gain on local/regional competition and incentives to be involved in knowledge transfer and experimentation to reduce costs; Allocative efficiency is based on the argument of better responsiveness to local preferences due to closeness to the people, facilitating collection of information and quality knowledge leading to immediate action. From the political point of view, elections can be used to reward or punish politicians when they fail to provide the public services efficiently. Decentralization promotes greater control over local politicians by putting them out of office in case of failure of being accountable. The literature also illustrates that local authorities become more accountable with the financial resources they receive because they will be judged on how they manage them and it affects directly their reputation and further re-election (Shaw and Qureshi, 1994 as cited in Tanzi, 1996). Hence, politicians will try to satisfy the median voter through improving democracy and accountability. By being more accountable to the people, it is argued to decrease lobbying by interest groups. Lobbying is thought to distort policy choice and allow waste of public funds (Barankay and Lockwood, 2007).

The expectation derived from the general arguments range from increments on economic growth, welfare and governance and decrements on regional disparities and

¹ The production of public goods at the lowest possible cost.

² The mix of goods and services that matches the goods and services desired by the local citizens.

poverty³. The theory has been neither completely vindicated nor disproved. However, caution have been put forward mainly in regard to fiscal decentralization. Fiscal decentralization refers to the provision of more authority on revenue collection through taxes and intergovernmental transfers to quasi-independent public bodies such as local governments. Among the counter-arguments, it is claimed that central government is weakened financially challenging the application of stabilization and redistributive policies; that regional redistribution is worsened due to competitive disadvantages among poorer regions and that administrative flaws and high level of corruption in local governments, make its successful application impossible ⁴. The latter issue is the concern of this paper.

An empirical research that contradicts the previous claims about local government inefficiency is an investigation carried out in Bolivia by Faguet (2000). Using an index of Unsatisfied Basic Needs and Local Government Effectiveness (LGE) constructed from a series of semi-structured interviews, he found evidence that contrarily to conventional wisdom, small-rural municipalities present highest LGE. In addition, he did not find strong correlation of LGE and urban-ness. Based on these findings, he proposes an incentive-based local leadership theory: In a context of mobile politicians and diversity of municipalities in regard to size and resources, corrupt politician will select wealthy municipalities leaving small-rural ones with honest politicians. Other empirical research that investigates local government performance is Moreno-Jaimes (2007). Taking as a case of study all Mexican local governments he investigated whether political competition made a difference in the performance of local governments. He did not find any support evidence of political incentives. Nevertheless, the results illustrate that indicators such as literacy rate, socioeconomic wealth and higher rate of voter participation seem to be the drivers of good governance. Based on these findings it is important to bear in mind that local government capacities and restrictions vary extensively. Rural local governments face different challenges and share large similarities in regard to socioeconomic status. For instance, rural local accountability falls into the general expectation of transparency, inclusion of vulnerable groups and good revenues management as the rest of local governments. However, entrepreneurship actions are limited to the resources available to the local governments. Hence, rural entrepreneurship might take place in different means than in other municipalities. Even the incentives of political elections might have different impact in a context where participation is not a major issue due to easy access to local authorities by citizens.

2.1 Background: The Mexican decentralization process

Mexico is a Federal country with 32 states, a Federal District and 2,457 municipalities by 2010. Mayors are elected democratically for three years without immediate reelection and councilors are elected via open-list proportional representation. The decentralization reform started in 1983 when municipalities were granted responsibility for potable water, drainage, sewage systems, public lighting, refuse

³ See Baskaran and Feld, 2009; Hammond and Tosun, 2011, Xie, Zue and Davoodi, 1999; Fukasaku and DeMello, 1998; Strumpf and Oberhalzer-bee, 2002.

⁴ See Prud'homme, 1995; Tanzi, 1996; Lessman, 2009.

collection, cemeteries, streets, public parks, public safety and slaughter house. By 1997 the federal government created a budget to provide with grants to States and municipalities (conditional grants)⁵ for specific areas besides the shared revenues (unconditional funds) they are entitled to. These grants increased the budget of local governments and even though, they are earmarked to certain expenditures, the local authorities take the decision of how to spend it and where. The amount to be granted depends on the State parliament which can select between two formulas stipulated in the law or its own.

The introduction of conditional funds in 1997 highly increased the local budget. Revenues from unconditional funds have always been the main source of revenue and the collection of own revenues from taxes and other contributions has slightly increased since the reform (see Figure 1).

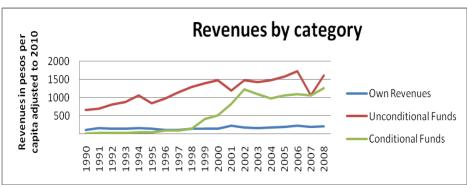


Figure 1: Revenues by category

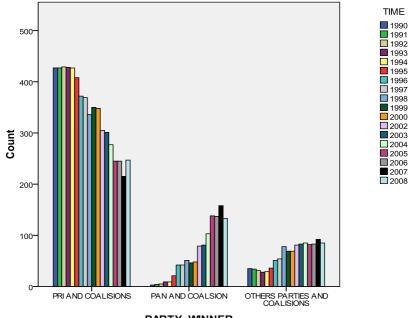
The fiscal decentralization reform described above took place during a changing political scenario. The Partido Revolucionario Institucional (PRI) was the ruling party in the country from 1929 until it handed over the presidency to the opposition party, Partido Acción Nacional (PAN) for first time in history in 2000. Progressively since 1990, PRI has been losing political control in all type of elections. Moreno-Jaime (2007) indicates that in 1990 almost 90% of the total population was under the mandate of local authorities coming from PRI membership and by 2001, more than half of this population was governed by opposition parties. In rural municipalities, this trend is following a slower pace (see Figure 2). From a sample of 465 rural municipalities and considering coalitions as part of the two strongest parties (PRI and PAN), PRI governed municipalities has changed from 91% in 1990, 74% in 2000 to

Source: Own calculation using SIMBAD database (Municipal System Database) by INEGI.

⁵ Unconditional grants are assigned trough the General Participatory Fund (GPF) via State. Federal government shares at least 20% to State and State shares, at least 20% to municipalities. Conditional grants includes FAISM (Fund for Social Infrastructure) and FORTAMUN (Fund for the strengthening of municipalities).

53% in 2008. PAN increased from an almost null presence in 1990 (3 parties only) to 10% in 2000 and 28% in 2008. Other left wing mainly PRD (Partido de la Revolución Democrática) and minority parties had more popularity in 1990 than the conservative party (PAN) starting with 7.3% in 1990, 14% in 2000 to 18% in 2008. After PAN won the presidency in 2000, its presence in rural local governments increased modestly.





PARTY_WINNER

Source: Own calculation using CIDAC-BANAMEX databases

3 Data and Methodology

The aim of this paper is to take advantage of the fiscal decentralization reform that took place in 1997 and uses longitudinal municipal data from the period 1990-2008 to investigate whether the reform has had any significant impact on the quality of governance measured by the coverage of Basic Public Services. In addition, it explores the influence of the decentralization reform on the local authority's behaviour by looking at changes in patterns of accountability and entrepreneurship after the reform measured by series of indicators (see Annex 5 for details on how variables are defined). The data used in this analysis comes from census (1990, 2000, and 2010) and counts (1995, 2005) collected by the Mexican National Institute of Statistics and Geography (INEGI), and surveys applied by different public and

academic institutions (1995, 2000, 2002, 2004 and 2009). Validation and model predictions are not the scope of this paper. The technique used was fixed effects design in Mixed Linear Model using SPSS. This technique allows for longitudinal database and it is flexible in exploring the covariance structure of the repeated measures.

3.1 Good governance at Local level: Before and after sceneries

The aim of this section is to describe the methodology used to investigate the changes on local government performance (dependent variable) before and after the 1997 reform using political and socio-demographic variables based on fixed-effect approach.

One measure that is related to local government performance is the Coverage of Basic Public Services (CBPS). The data available from INEGI (census and counts) is the coverage of sewage and water services. The coverage of the service is interpreted as the ratio between the number of houses with water/sewage service and the total number of houses in the municipality in the same census or population count⁶. The indicator is measured as follow: changes in percentage points in the average coverage of water and sewage service during a period of five years (one census and one count.). In order to determine yearly percentage, a progressive increment between one census and one count was calculated⁷. The periods are divided into pre-reform period from 1990 to 1997 and post-reform period from 1998 to 2008 (See Annex 1, Table A for descriptive statistics).

Next, the model used two set of explanatory variables, political and sociodemographic. The political variables include juxtaposition, electoral completion and political party in order to explore how the political scenario changed. The sociodemographic variables consist of territory and population characteristics.

The variable called Juxtaposition was introduced by Moreno-Jaimes (2007) in a study of Mexican local governments. The variable was used to identify those municipalities where the party membership of a local mayor and the governor come from different parties. His study was based on all local governments in Mexico from 1990 to 2000. He uses the same dataset but different approach. His hypothesis was that juxtaposed local governments will have more limited access to State government funds, and therefore, lower coverage of services. He did not find any differences in service coverage in juxtaposed municipalities. In the context of this study, the same definition of juxtaposition is used. There is juxtaposition when mayors are from different political membership than the governor. The variable takes a value of 1. In contrast, if the mayor is from the same political party than the governor, it takes the value of 0. However, in this study the hypothesis is that municipalities juxtaposed before the reform had limited access to resources allocated to basic public services by the State due to political rivalry than those from the same party. This will lead to lower mean scores in CBPS in the pre-reform and, either no-mean difference or

⁶ The information for these variables is listed on the Census of Population and Housing 1990, 2000 and 2010 and Count of Population and housing 1995 and 2005 from INEGI.

⁷ The reason to include a yearly measure is because the census and counts does not match the time of local administrations. By including them, it allows the inclusion of three years span in the analysis to each local administration.

higher score for juxtaposed municipalities after the reform due to incentives to perform better in hostile political environment but entitle to allocate public services budget⁸ (see Annex 1, Table B for descriptive statistics). Hence, the main difference with respect to his paper is that Moreno-Jaimes explored whether juxtaposed municipalities were more prone to budget restriction. He considers the intergovernmental transfers as another explanatory variable. Differently, this paper looks for insights of positive incentives with the introduction of the 1997 reform in regard to CBPS even in case of juxtaposition considering the fact of increments in local budgets.

Following Moreno-Jaimes study, political competition is included in order to observe its influence in rural settings. The indicator of competition in local elections (electoral competition) is measured as 1 minus the difference in the share of votes obtained by the two strongest parties. Therefore, a high index denotes high levels of competition and a low index the opposite. For instance, Xicoténcatl, my home town, in 2001 had local election. For first time in the history of the town the opposite party won the election. The total votes were 11,405, and the share of votes by party was PAN (opposite party) 5,657, PRI 5,447 and the rest accounted for other parties and canceled votes. The difference between the two strongest parties was 210 votes. This represents 0.02 or 2% difference in percentage. Hence, the index results in .98 (1-0.02) or 98%. This was a highly competitive election (see Annex 1, Table B for descriptive statistics).

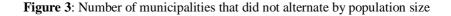
The third set of electoral variables included is an identifier of groups of parties that were governing each year⁹. The two strongest parties also include coalitions.¹⁰ Besides, three socio-demographic variables are included: territory, measured in squared kilometers, population by 1000 inhabitants and population density. The reason to include these variables is because the cost of the basic public services will be higher for municipalities with larger territory and population density. Therefore, the coverage of services might be lower for them. Moreover, even though this research sampled only small municipalities with rural characteristics, the number of inhabitants is politically important for state and district elections. In addition, municipalities with low population are the most behind in the alternation trend in local elections. By alternation this study means that any other party or coalition but PRI and its coalition have won the local election at least once. From the 465 sampled municipalities, the percentage of municipalities with at least one alternation changed from 8% in 1990 to 19% in 1997 to 62% in 2008. However, if divided into population

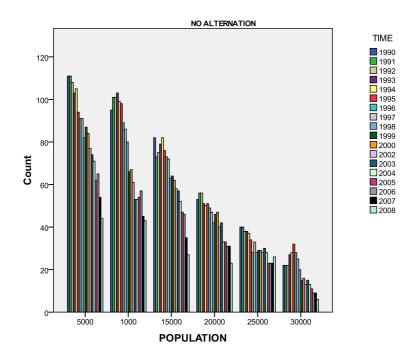
⁸ This dummy variable follows the same mechanism than the grouping of parties. If a party from a coalition, example PRI plus PANAL (Partido Nueva Alianza) is governing the State, a municipality from PRI and/or the same coalition is included as NO JUXTAPOSED. PANAL in its own is considered as JUXPOSED.

⁹ Partido Revolucionario Institucional (central-left wing) and coalitions (PRI+C); Partido Acción Nacional (right wing) and coalitions (PAN+C); Others + Coalitions (OTHERS+C).

¹⁰ In some cases it lasted more than 3 years due to special circumstances. Competitive index and Party winner are set in the years that local governments were governing in order to match their effort for increasing local services. Elections in Mexico are not held the same month and the starting of the administrative period neither. However, in most cases the elections are held around 6 months before the start of the administrative period and at middle or later months of the year. Hence, if the election was held in 1990, the competitive index as well as the winner of that election is given to the subsequent three years (1991, 1992, and 1993).

size, most municipalities with small population size have always been governed by the hegemonic party (PRI) (see Figure 3). Hence, distribution of resources might have been threatened even for municipalities from the same party where political power was strongly rooted by the hegemonic party.





Source: Own calculation using CIDAC databases

In addition to the main effect analysis of the previous variables, an interaction between electoral competition and juxtaposition is included. The aim is to understand whether any changes may be attributed to the implementation of the 1997 reform, which gives financial decision making to local governments; or might be a consequence of the progressive uprising of political competition, which according to the theory promotes accountability.

Although the interaction must be read in a contextual mean, the logic that follows is: If the higher the political competition, the average CBPS is lower for juxtaposed municipalities than for no-juxtaposed before 1997 and it changes to higher average CBPS after the reform, it suggests that the reform played an important role on promoting better governance. The reason can range from deprivation of favoritism to decision making empowerment. In contrary, if the higher the political competition, the average CBPS is higher for juxtaposed municipalities than for no-juxtaposed in both before and after the reform, it might imply that electoral competition is the main incentive of good governance rather than the reform.

It is important to notice that the regressions include municipal fixed-effects, which allow controlling for any time-invariant characteristics of the municipalities that might have had an effect on CBPS over the period under analysis (e.g. State development). Moreover, each analysis (before and after) is estimated from a sample restricted to years before 1997called BEFORE-REFORM and after 1997 called AFTER REFORM. It estimates how political and demographic characteristics influence changes in CBPS in two different sceneries for local authorities: in the first scenery decision making in budget allocation is made at higher government levels (before reform) and in the second one, local authorities are empowered to a certain degree of decision making (after reform).

3.2 About internal changes in local governments: accountability and entrepreneurship

This section outline the methodology used to address the question whether fiscal decentralization has been associated with changes in patterns of accountability and entrepreneurship. Firstly, it is important to define these concepts.

Accountability is an ambiguous concept. Kluvers and Tippett (2010, 47) paraphrased it as '...broadly speaking, the process via which a person or group can be held to account for their conduct' (Glynn and Murphy, 1996). The public sector is an interactive system with the problem of multiple principals. Hence, accountability can be seen from the managerial point of view focusing on the daily operations of the organization and also from the political accountability side which involves issues of democracy and trust (Boadbent and Laughin, 2003 as cited in Kluvers and Tippett 2010). Hence, the financial autonomy derived from the fiscal decentralization policy should produce internal changes in accountability for both internal and external stakeholders as a mean of allocative efficiency. The concept of accountability leads to scrutiny of the level of openness and transparency in the local governments, the mechanism encouraging citizens' participation and involvement in decision making, its legal compliance toward financial resources, and its commitment to equality in terms of political affiliation and vulnerable status (inclusion). Hence, it deals more with allocation efficiency. This concept is easily applied to the rural setting.

In regard to entrepreneurship, it is a concept borrowed from the private sector. The public administration literature define public entrepreneurs as 'individuals who undertake purposeful activity to initiate, maintain, or aggrandize one or more public sector organizations' (Ramamurti, 1986, 143 as cited in Zerbinati and Souitaris, 2005). More recent studies have moved from a definition of economic maximization toward innovation and creativity. Zerbinati and Souitaris (2005) adopted a concept from Shane and Venkataraman (2000). They explain that Shane and Venkataraman (2000)'s definition implies that entrepreneurship in the public sector involves innovative action rather than just wealth creation. The concept is still not as straightforward. Moreover, it is more difficult to understand its expectation in the context of rural municipalities where social capital and resources are scarce.

After a series of interviews¹¹ in a pilot study carried out by the author, three types of entrepreneurial actions were identified (see Annex 4): **Grant application-facilitation-** Local government can assist and participate actively with local groups for being eligible for funds granted by the state or federal government directly for specific sector; **Social-cooperation management-** The use of social cooperation and compulsory social services to minimize costs for local projects; **Win-win Negotiation-** Local government can negotiate with groups and communities for sharing costs for focalized projects.

Hence, entrepreneurship in rural local governments in this study is defined as the use of cooperation with private, public and social bodies and the ability to negotiate with the community in order to reduce cost/ increase revenues. It deals more with productive efficiency rather than allocation.

It is a challenge to measure such abstract concepts. Its accuracy is debatable, but it is the only available mean of understanding the general changes in a summarized way. Moreover, it is useful in order to explore general trends as well as identify specific cases¹². For the use of this paper and following the definition mentioned above, two indexes were constructed. The data comes from five local government surveys carried out in 1995, 2000, 2002, 2004 and 2009 and financial data from 1990 to 2009. Due to missing data, the entrepreneurship index only includes 2000, 2002 and 2009 surveys. Each variable have a maximum score of 2 and a minimum of 0. Some variables take an intermediate score of 1. The score of the indicator represents whether the municipalities have good, moderate or poor performance. The variables with three levels come from financial data and few from the surveys as liker scale variables. In most of the cases, the limits of each level were set according to the sample studied. The highest mean score of a variable over time is set as the upper limit and the average of all mean scores as the limit between moderate and low. This adds an important element to the analysis; it is based on an exclusive analysis for rural local governments, therefore, it shows a more realistic approach of what can be achieved and, how good or bad rural local governments are performing (See annex 5). The variables selection and measurement were largely benefited from the work of Carrera, A., Coronilla, R. and Navarro, A. of the Association of researches network in Mexican Local Governments.

The indexes show that local authorities have been progressing over time in both accountability and entrepreneurship (See Annex 1, Tables C and D). These changes are significant (Annex 6, table B and D).

¹¹ The interviews were held to 5 main local actors in a local government in Mexico. They represented elected members including two mayors and one local council, a managerial level and an external stakeholder.

¹² In further analysis, the construction of these indexes will be used to identify municipalities with strong changes in accountability and entrepreneurial actions in order to do an in-depth qualitative analysis.

4 Empirical Analysis

In regard to local government performance, the average mean of CBPS at the starting of the analysis was 45.18% and the average score is lower by 1.26 points in CBPS before the reform. The rate of change was higher in the pre-period by .32 with an average of 2.04 points growth every year before the reform. It slowed down to an average of 1.72 after the reform (Annex 2, table A)¹³. Hence, there were no major changes in the provision of services after the introduction of the reform. This is not a concern itself because the reform does not involve increments in resources but a change of hands of the resources and particularly, efficiency in the allocation of resources within the country by avoiding vertical and horizontal imbalance¹⁴.

The estimation results for both models are presented in Table 1. Both analyses present roughly the same pattern of statistical significant determinants of CBPS but with changes in the estimated magnitudes of their effects. Socio-demographic variables were not statistically significant once controlling for the other variables in the analysis. Only the political variables were correlated and significant (See Annex 3, table A and B for the model with statistically significant political competition variable). Electoral competition variable has the expected results. The more competitive the local election, the more coverage of service a municipality has. Parties from PAN had slightly higher mean score than PRI taking OTHERS+C as a comparative case. However, the dummy variable PRI is not significant in both cases. This might be due to personal characteristics from the PRI members

Contrarily to the stated hypothesis, the juxtaposition dummy variable shows that the mean of municipalities with governors from the same party, in overall, have slightly lower means than those from different parties in both before and after the reform. It would be expected that no juxtaposed municipalities will be benefited with resources by the state government from the same party. This logic appears with the introduction of an interaction of Juxtaposition by Political Competition. This interaction turns down the significant level of electoral competition but not the political dummy variables (see table 1). Hence, the interaction did not follow the proposed logic but illustrates other matter. It shows that in higher political competition, municipalities that are from the same political party than the governor in average have higher mean than those juxtaposed. One explanation is that a competitive election for non-juxtaposed municipalities means a probability to lose the next election if the local

¹³ The model accounts for covariance-structure with auto-regression. This allows the correlation to decrease over time. The ceiling effect is not still highly a concern The curve fit for whole database present the following results: Linear (R Square= .678, DF= 3253, Sig .000), Quadratic (R Square= .685, DF= 7902, Sig .000), Cubic (R Square= .685, DF= 7901, sig. 000). By period: BEFORE REFORM, Linear (R Square= .864, DF. 3253, Sig .000), Quadratic (R Square= .864, DF. 3253, Sig .000), Quadratic (R Square= .869, DF= 3252, Sig .000), Cubic (R Square= .870, DF= 3251, Sig .000); AFTER REFORM, Linear (R Square= .660, DF= 4648, Sig .000), Quadratic (R Square= .670, DF 4647, Sig .000), Cubic (R Square= .670, DF= 4646, Sig .000). Because they show not high difference, the simplest model was chosen.

¹⁴ Vertical imbalance occurs due to high levels of services and low level of revenues to afford such services and horizontal imbalances results from difference in wealth and development within the country as well as political manoeuvres

authorities do not improve their political image during the local administration. Hence, resources might be redirected to these municipalities. In other scenery where the same municipality wins without high levels of competition, the State party (headed by the governor) will not prioritize redirection of resources. However, once a rural municipality is in hands of the opposition, the political weight of the municipality might play a role in the redirection of resources by the State incumbent. As stated before, alternation within rural local governments is emerging slowly. The alternation appeared first in developed cities. Hence, an explanation of lower mean on non-juxtaposed municipalities is that State incumbents might prioritize between redirecting resources to developed municipalities rather than rural. Hence, higher CBPS in juxtaposed municipalities might be the sign of good performance by opposition parties when the political weight of no-juxtaposed municipalities is not substantial.

 Table 1: Results of fixed effects analysis of political variables BEFORE

 REFORM and AFTER REFORM

Parameters	Estin	Estimates		
	BEFORE	AFTER		
	REFORM	REFORM		
Intercept	61.268***	68.385***		
	(1.150)	(1.145)		
POLITICAL COMPETITION	.889	.437179		
	(.527)	.735		
YUXTAPOSITION $= 0$	-2.386***	-4.041***		
	(.487)	(.763)		
PRI+C	.187	.089		
	(.145)	(.171)		
PAN+C	.522**	.520**		
	(.176)	(.203)		
[YUXTAPOSITION=0] * POLITICAL	2.044***	3.780***		
COMPETITION	(.549)	(.856)		
Pseudo R-squared	33.90	26.2		
N-1	464	464		

Note: Dependent variable CBPS. Standard errors are in parentheses. McFadden's pseudo R squared was used. The result is a measure of the improvement in fit of the model due to the additions of independent variables.

N-1(464) ***p<.001 **<.01 *p<.05

Hence, the virtues of political competition and elections as a mechanism to keep local authorities accountable are partially challenged. The general results support Moreno-Jaimes' (2007) findings. However, it adds the wisdom of the underneath effects of political competition in rural municipalities, namely its importance as a warning sign of losing power in non juxtaposed settings. In addition, it also suggests that virtues of political competition can be reached if there is not political intervention from the State government. However, at this point, it can not be claimed either that political competition or the introduction of the reform made a difference in performance.

In regard to changes within local government structure, both indexes accountability and entrepreneurship are highly and significantly correlated to Coverage of Basic Public Services. Therefore, it is expected to find similar results than in the previous analysis (see Annex 6, table A and C). The estimates results are presented in Table 2 and 3.

Electoral competition is not significant in the accountability index neither as a main effect nor as an interaction with juxtaposition. In entrepreneurial index (see table 3), it turns significant if there is an interaction with juxtaposition as in the previous analysis. The interpretation in this case is that municipalities not juxtaposed and with high level of political competition in the previous election will be encouraged to perform better in order to avoid losing the next election. Interestingly, the financial arrangements show that unconditional funds are negatively related to accountability index and insignificant in the entrepreneurial index while the conditional funds are highly positively correlated on both. The negative sign in the accountability index can be thought to be a consequence of the variable public work investment. However, the exclusion of individual financial compliance variables in the index does not change the negative correlation, but exclusion of both turns it insignificant. This suggests that increments in unconditional funds not only foster waste of resources in administrative spending but also discourage investment in public work investment. On the other hand, a highly positive correlation of conditional funds and entrepreneurship suggest that increments on conditional funds encourage cooperation, which might reduce cost, and fiscal effort, which might lead to investment.

Hence, the analysis suggests that financial arrangements play an important role in motivating local authorities to act accountable and entrepreneurial and that political competition acts as an incentive only in case it is used as a warning sign in a municipality of being losing political power that treats further elections.

Parameters	Estimates	
Intercept	.152***	(.046)
ZELECTORAL COMPETITION	037	(.047)
ZUNCONDITIONAL FUNDS	192***	(.032)
ZCONDITIONAL FUNDS	.247***	(.026)
YUXTAPOSITION=0	251***	(.054)
[YUXTAPOSITION=0] * ZELECTORAL COMPETITION	.078	(.055)
R-squared	.055	

Table 2: Results of fixed effect analysis of accountability index

Note: Dependent variable CBPS. Standard errors are in parentheses. McFadden's pseudo R squared was used. The result is a measure of the improvement in fit of the model due to the additions of independent variables.

N-1 (264) ***p<.001 **<.01 *p<.05

Table 3: Results of fixed effect analysis of entrepreneurship index

Parameters	Estimates	
Intercept	.251***	(.068)
ZCONDITIONAL FUNDS	.229***	(.041)
ZELECTORAL COMPETITION	130	(.077)
YUXTAPOSITION = 0	296***	(.081)
[YUXTAPOSITION=0] * ZELECTORAL COMPETITION	.203*	(.088)
R-squared	.021	

Note: Dependent variable CBPS. Standard errors are in parentheses. McFadden's pseudo R squared was used. The result is a measure of the improvement in fit of the model due to the additions of independent variables.

N-1 (207) ***p<.001 **<.01 *p<.05

5. Conclusion

This paper shade lights on the incentives generated by fiscal decentralization policies and the political environment in rural municipalities in Mexico. The results did not find any significant differences in patterns of growth in CBPS before and after the reform. Allocation of resources due to socio-demographic characteristics did not influence on the CBPS for this group of municipalities. Political competition is only relevant when the electoral competition is high and the municipalities are from the same party than the governor. Financial incentives via conditional funds are found to be important in promoting accountability and entrepreneurship while unconditional funds have a negative effect in accountability and no effect in entrepreneurship.

This research supports the arguments of the decentralization theory in regard to incentives to behave more accountable when financial resources are passed down. However, it does not support the argument of autonomy as a driver of good performance in rural municipalities and pure political election as an incentive to be neither accountable nor entrepreneurial. It can be concluded that Mexican municipalities have been engaged in positive changes at the rural local structure that promotes level of accountability and entrepreneurship. Hence, the fiscal decentralization policy has been beneficial.

The type of financial arrangement plays a main role in the sense of direction for performance at the local government. Conditional funds should be highly considered. But, there are different arrangements to share funds. Do they make a difference in the entrepreneurial behavior of rural local authorities? This is a further research question...

References

Barankay, I., Lockwood, B. (2007) Decentralization and the productive efficiency of government: Evidence from Swiss cantons. *Journal of Public Economics*. 91,5-6,1197-218.

Baskaran, T., Feld, L.P. (2009) *Fiscal Decentralization and Economic Growth in OECD Countries: Is there a Relationship*, CESifo Working Paper Series No. 2721. Available at SSRN: http://ssrn.com/abstract=1441152.

Carrera, A., Coronilla, R., Navarro, A. (2010) Indice de Desarrollo Institucional y Sustentabilidd Municipal. Working paper. Preliminary Version.

Ezcurra, R., Pascual, P. (2008) Fiscal decentralization and regional disparities: evidence from several European Union countries. *Environment and Planning*. 40,11,85-201.

Faguet, J.P. (2000) Decentralization and local government performance: improving local service provision in Bolivia. *Revista de economía del Rosario.* 3, 1, 127-176.

Freitag, M., Vatter, A. (2008) Decentralization and Fiscal Discipline in Sub-

national Governments: Evidence from the Swiss Federal System. *Publius: The journal of Federalism.* 38, 2, 272-294.

Fukasaku, K., and DeMello, L. (1998) *Democracy,Decentralization and Deficits in Latin America*. Inter American Development Bank/ Development of the Organization for Economic Cooperation and Development, Paris.

Hammond, G., and Tosun, M. (2011) The impact of local decentralization on Economic Growth: Evidence from US counties, Journal of Regional Science, 51,1, 47-64.

Kluvers, R. and Tippett, J. (2010) Mechanism of accountability in local governments: an explanatory study. *International Journal of Business and Management*. 5,7, 46-53.

Lessmann, C. (2009) Fiscal decentralization and regional disparity: evidence from cross-section and panel data. *Environment and Planning A. 41*, 10, 2455-2473.

Litvack, J. and Seddon, J. (1999) *Decentralization briefing notes*, World Bank Institute Working Paper, World Bank, Washington

Moreno-Jaimes, C. (2007) Do competitive Election produce Better-Quality Governance? Evidence from Mexican Municipalities, 1990-2000. *Latin American Research Review.* 42, 2, 136-153.

Oates, W. (1972) Fiscal Federalism. Harcourt Brace Jovanochi, New York

Oates, W. (1999) An essay of fiscal Federalism. *Journal of Economic literature*, 37,3, 1120-49.

Prud'homme, R. (1995) On the dangers of Decentralization. *The World Bank Researcher Observer.* 10, 2, 201-220

Rodriguez-Pose, A., Ezcurra, R. (2009) Does decentralization matters for regional disparities? A cross-country analysis. *Journal of Economic Geography*, 10,5, 619-644.

Strumpf, K.S., Oberholzer-Gee, F. 2002 Endogenous Policy Decentralization: Testing the Central Tenet of Economic Federalism. Journal of Political Economy. 110, 1, 1-36

Tanzi, V. (1996). Fiscal federalism and decentralization: A review of some efficiency and Macroeconomic aspects in edited volume, in Bruno, M, Pleskovic, B (eds), *Annual World Bank Conference on Development Economics*, World Bank, Washington, 295-316.

Tiebout, C.M. (1956) A Pure Theory of Local Expenditures. *The Journal of Political Economy*, 64,5, 416-424.

Xie, D., Zou, H. and Davoodi, H. (1999) Fiscal Decentralization and Economic Growth in the United States. *Journal of Urban Economics*. 45,2, 228-239.

Zerbinati, S. and Souitaris, V. (2005) Entrepreneurship in the public sector: a framework of analysis in European local governments. *Entrepreneurship and Regional Development*. 17, 1, 43-64.

Foreign Direct Investment through the imitation procedure: The case of China

Polyxeni Kechagia¹ and Theodore Metaxas²

¹ PhD student, Department of Economics, University of Thessaly, Volos, polyxeny84@yahoo.gr

² Lecturer, Department of Economics, University of Thessaly, Volos metaxas@uth.gr

Abstract. Over the past decades great changes have occurred worldwide regarding the foreign direct investment (F.D.I.). However, certain developing countries have managed to gain more F.D.I. compared to other developing countries via skills acquisition, competition, exports and imitation. These spillovers channels are used to transfer technological knowledge among firms, while the imitation is used by many multinational enterprises and it is listed among the most important spillovers channels. China is the largest recipient of F.D.I. inflow and the country's development affects the policies and the growth of other developing countries. In the present essay a literature review is conducted so as to investigate whether the case of China could be an example to be imitated regarding the F.D.I. inflow by other developing countries. The contribution of the essay to the existing literature refers to the fact that it is not limited on the conditions under which the imitation procedure is performed. On the contrary, it focuses on the factors that affect the attitude of the developing countries towards imitating China and on the categorization of these countries into regions, rather than comparing China with a certain country. The main finings of the research suggest that under certain circumstances, developing countries have the ability to attract F.D.I. through imitating China. Such circumstances include the development of the absorptive capacity and the sufficiency of the linkages between China and the developing countries. Finally, it is suggested that developing countries imitate China in order to take advantage of the benefits deriving from F.D.I.

Keywords: *imitation, China, developing countries, spillover channels*

1 Introduction

Over the last years Foreign Direct Investment (F.D.I.) inflow has increased significantly. Countries such as China, India, Brazil and Mexico have become major recipients of F.D.I. inflows. The case of China has been selected because the country has been the greatest recipient of F.D.I. compared to other developing countries and it is very interesting because of the country's great variety among the regions, of the large geographical area it covers, of the great economic growth and of the country's large market (Qi et al., 2009:18-35). The purpose of the present study is to examine whether neighbouring to China and developing countries can imitate China's attitude towards F.D.I. so as to attract foreign capitals.

Hence, according to Gorg and Greenway (2004) who studied theoretically and econometrically the productivity spillovers, domestic firms gain positive spillovers, through imitation, skills acquisition, competition and exports. Also, they proposed that domestic firms set operations abroad so as to gain higher rate of return compared to the one they would have gained from the home country investment (Gorg & Greenaway, 2004: 171–197).

So far there are only few studies worldwide connecting the imitation procedure and the F.D.I. (Glass & Saggi, 1999, 2002; Glass & Wu, 2007, Gorg & Greenway, 2004) and most of them focused on the imitation cost (Mansfield et al, 1981; Romer, 1990). The present study, unlike past researches (Gorg & Greenway, 2004), does not solely examine the conditions that contribute to the spillovers effects and to the imitation procedure. Therefore, the factors that affect the developing countries' attitude towards imitating China so as to attract F.D.I., such as China, are presented. Also, the developing countries that are possible to imitate China are categorized based on the region they are located in, that is to say into Asian, European and Latin American countries. Therefore, unlike past studied (Adams, 2003; Fung et. al., 2009), the case of China is not compared to a certain country, but to the countries of a region.

The paper is organized as following: Section 1 presents the spillovers channels, section 2 analyses the imitation procedure for neighbouring to China and other developing countries and section 3 concludes with the main findings and suggestions.

1.1 Spillovers Channels

Spillovers refer to the transfer of technological knowledge and other resources among enterprises that have no contractual agreement (Meyer, 2004: 259–276). According to Caves (1974), foreign enterprises augment the technological spillovers effects on the host country via competition and transactions. Also, foreign enterprises demonstrate the advantages of the technology they used to the host enterprises, which leads to an imitation procedure. Finally, it has proven that there is a positive relationship between the labour productivity of the local enterprises and the share of the employment for the foreign enterprises in a sample of 22 Australian firms (Caves, 1974:176-193). Other studies reached to similar conclusions and confirmed Caves' findings. Hence, Globerman (1979) proved that there is a positive relationship between labour productivity and measures for foreign enterprises in Canadian firms (Globerman, 1979:42–56).

However, other studies reached to the conclusion that there are no technological spillovers effects from foreign to local firms. Hence, Haddad and Harrisson (1993) did not observe any changes of the productivity of local enterprises of Morocco

deriving from foreign enterprises (Haddad & Harrisson, 1993: 51–74). Moreover, in Venezuela the productivity of the local enterprises was reduced because of the foreign ownership (Aitken & Harrisson, 1999: 605–618).

Multinational enterprises perform F.D.I. so as to gain a higher rate of return compared to the rate of an investment in a domestic firm because of the technological advantage. There are numerous channels through which F.D.I. influence the countries that receive foreign capitals. The present study is based on the four categories of these channels proposed by Gorg & Greenway (2004): skills acquisition, competition, exports and imitation.

1.1.1. Skills Acquisition

Skills acquisition or investment on human capital also contributes to the adoption of the technology advances. Acquiring firms demand skilled employees, even if the wages are relatively low. Higher productivity can be realized because of the movement of employees through direct spillovers to new employees or through the knowledge employees who move to the new firm already have (Fosfuri et. al., 2001: 205–222).

Moreover, multinational enterprises have made considerable efforts in developing educational programs for the local workforce and they offer more training to workers and managers rather than the local businesses do (Motta & Ronde, 2000). The possibility of diffusion is maximized when employees of subsidiaries move to the parent company or establish their own business. The possibility of diffusion of human capital in Mexico, Venezuela and the U.S. has been examined in a previous study, which reached to the conclusion that F.D.I. improves the human capital of the domestic employees of foreign subsidiaries. However, there is no evidence of diffusion of human capital to employees of domestic enterprises (Aitken et al, 1996: 345-371).

1.1.2. Competition

Competition is another important channel for spillovers. Foreign firms have to produce in competition with the local ones. Also, local firms are forced to use the existing technology more efficiently. Competition is considered one the main sources of gain because greater competition leads to the reduction of X – inefficiency and it increases the speed of adoption for the technology advances (Wang & Blomström, 1992)

The entry of a foreign company initially increases the competition in the domestic industry, which boosts domestic firms to adopt new technologies or to reduce X-inefficiency, even if there are no gains in terms of imitation. This mechanism is similar to the gains associated with increased trade and it is often characterised as one of the most important benefits of the F.D.I.. Of course, if the entry of foreign enterprises displaces some domestic companies that are unable to compete and ultimately leads to increased concentration and market imperfections, then the impact on competition from F.D.I. may actually harm the economy of the host country (Glass & Saggi, 2001: 67- 86; Wang & Blomstrom, 1992).

Also, it has been taken under consideration the fact that the competition between foreign and local firms is affected by cultural factors and that host countries have poor infrastructures in social, economical and political level (Hymer, 1960). Moreover, it is

possible that foreign firms do not have direct access to local knowledge. Finally, it has been proposed that local firms take advantage of the benefits that the governmental policies of the host countries provide to them and that they have created and participate in a social network that prevents the competition with foreign enterprises (Makino & Delios, 1996: 905–927; Lu & Xu, 2006: 427–448).

1.1.3. Exports

There are several studies examining the exports as a spillovers channel. Aitken et al (1997) studied the relationship between F.D.I. and exports in Mexico. They noted that the probability of a local firm to export is positively correlated with proximity to subsidiaries of multinationals, but not with the general export activity. They concluded the foreign-owned firms are a natural conduit for information about foreign markets and technology and a natural channel through which domestic firms can distribute their products (Aitken et al, 1997: 103-132). Also, exports include the cost of developing distribution networks and infrastructures. Domestic enterprises can apply exports, via imitation or collaboration, so as to reduce the cost of operating in the foreign market, thus enhancing the productivity of domestic enterprises (Bernard & Jensen 1999:1-25). Finally, exports have been enhanced by the globalization process which leads to the convergence of the developing countries' economies. Moreover, trade barriers have been limited and new markets have been created to the developed and developing countries (Dunning, 1995).

1.1.4. Imitation

There are two activities composing the production sectors of a country's economy; the innovation and the imitation procedure. Every firm has to learn the production technique for each product. This can be achieved by innovation, in case the product is brand new, or by imitation, in case the product is already produced and distributed. Either way, the enterprises have to dispose a certain amount of resources so as to produce in a stable returns – to –scale production. Finally, imitation is considered the only channel through which the technological knowledge is diffused among countries worldwide (Grossman & Helpman, 1991: 1214–1229; Krugman et.al. 1979:253-266, Segerstrom 1990:1077-1091). Therefore, the present study focuses on the imitation procedure regarding F.D.I. among developing countries.

Imitation is the procedure through which mechanisms and processes are transmitted. Reverse engineering is the most commonly mechanism used in order to transfer technology from the acquiring to the acquired firm (Das, 1987: 171-182; Wang & Blomström, 1992). This mechanism allows firms to imitate not only the easier manufactures and processes, but also the managerial and organizational innovations. So, spillovers for the acquired firm derive from the technological upgrades and they lead to more productivity for the local firms (Wang & Blomström, 1992).

The imitation effect occurs in case there is a technological gap between the local and the foreign firm. The imitation effect is measured as the share of affiliate enterprises' R&D to the total R&D of a productive sector. Foreign firms' R&D affects indirectly the export capacity since local enterprises are enhanced to produce at high technological level and they become more productive so as to survive in foreign markets. Hence, it is more possible for foreign firms that use technology at high level to be imitated compared to other enterprises that do not use recent technological advances (Franco & Sasidharan, 2010:270-288). Also, the domestic market in developing countries is usually small; therefore it is important for the firms of these countries to sell to the markets of the developed countries so as to develop as well. However, high cost and demand factors make it block the developing countries' investment in R&D. Also, it is suggested that the imitation procedure leads to the reduction of the wage gap worldwide and developed countries have more incentives to innovate. However, when the transaction costs are low worldwide then it is possible that the imitation will affect negatively the wage inequalities and the development worldwide (Kind, 2004:47-67).

Foreign enterprises use recent technological advances so as to enter to the new market and introduce these advances to the local firms of every sector. Hence, local firms imitate the way foreign enterprises operate and this gives them the opportunity to enhance the productivity. Apart from imitation, local firms are also affected by the additional competition foreign firms create. This competition makes local enterprises operate more efficiently and they become more innovating so as to preserve their market position. This imitation effect, as described above is more probable to occur at the intra-industry level (Fosfuri et al. 2001: 205–222).

However, when developing countries adopt new technology, it is necessary to replace the old capital with new one. Therefore, F.D.I. is considered as an engine that promotes rapidly the technological development. The know-how is diffused from developed to developing countries via F.D.I., facing though various problems. Restrictions and barriers to the inflow investment, protection policies of the domestic firms ect. limit the technology transferred to developing countries (Institute of Economic Affairs, 2005). Also, imitation is only profitable for developing countries when the trade is sufficiently liberalized because the imitation procedure has positive effects when the access to the markets of the developed countries is cheap. This means that there is a positive relationship between trade liberalization and the imitation procedure (Kind, 2004:47-67).

Also, over the past 15 years developing countries have strengthened the intellectual property rights (IPRs.), thus restraining the imitation procedure. Developing countries have limited access to new technologies applied in developed countries because of the foreign enterprises' monopoly. However, it is difficult to determine which way the strengthening of IPRs affects the imitation procedure and the technology diffusion between developed and developing countries (Ivus, 2011:201-226).

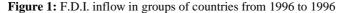
Especially over the past years, because of the recent financial crisis and the problems deriving from it, the need for imitation through F.D.I. has become more intense. That's because F.D.I. are used mostly for long – term production plans, so they are characterised by stability, they are more resilient compared to other types of capital flows and, finally, the cost of F.D.I. is lower compared to the that of external borrowing and portfolio equity financing (Yan, 2007: 104 - 120). Besides, apart from the recent financial crisis, the Asian one has led to a discontinuity in the F.D.I. flows' story in the region and has caused the reduction of portfolio and short – term investments, rending the imitation procedure more important in facing this crisis (Hill & Athukorala, 1998:23-50)

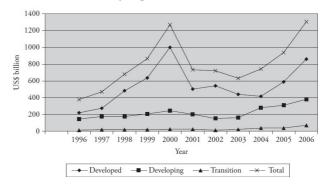
In other words, an imitation procedure is considered successful only in case that the products used as input are intermediate and that these intermediate products are produced in developing countries. Thus, if the transaction costs are high, then it is

expensive for the developing countries to imitate the developed ones. Also, the imitation procedure requires fixed investment in R&D and it is considered profitable only when the demand is sufficiently high (Kind, 2004:47-67).

Over the past three decades the developing countries, as well as the transition economies have managed to receive an increasing amount of F.D.I. inflow. Therefore, in 1990 the amount of F.D.I. inflows for the developing countries was estimated at 17,8%; however in 2004 it reached at 36,61% and in 2005 at 35%. The F.D.I. inflows to the developing countries reached at \$916 billion in 2005, increased by 27% compared to the inflows in 2004. These inflows were distributed in several sub – regions of the developing countries and they led these countries into economic development. As for the transition economies, the F.D.I. inflows rose from 0,04% in 1990 to 5,57% and 4,33% in 2004 and 2005 respectively. Developed countries, mostly due to several mergers and acquisition (M&As) which led them to higher development rates, received \$168 billion in 2000, a year that the total inflows reached their peak (\$1,4 trillion). (Cevis & Camurdan, 2009:210-223).

It is obvious that the capitals inflows reduced in 2000. In particular, the share of the developing countries reduced from 38,4% in 1996 at almost 18% in 2000 (Figure 1). This reduction has caused serious concern because some of the developing countries did not manage to attract the F.D.I. inflow, although they were in great need; thus it was suggested that these countries should be further liberalized (Nanda, 2009:26-47).





Source: Nanda (2009)

In our study the case of China is examined because it is the greatest recipient of F.D.I. inflows nowadays. The methodology chosen for the present study is the literature review of secondary data deriving from international journals, books, research papers ect. In particular, the case of China is chosen because over the past three decades the country managed to become the second recipient of F.D.I. inflow worldwide, after the U.S.A., and the first recipient among the developing countries. The country's F.D.I. inflow rhythm rise significantly from 1992 up until today because China became more open – oriented to F.D.I. and repealed some special regimes that block the country's growth. The F.D.I. inflow and the exports and imports in China from 1985 to 2006 are presented in the following figure (Fu, 2008:89-110).

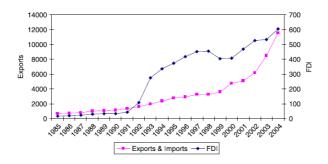


Figure 2: F.D.I. inflow, imports and exports in China from 1985 to 2006*

Source: Statistical Yearbook of China, published in Fu (2008), * in \$ 100 million

The reforms chosen by the Chinese government in order to attract most F.D.I. inflow refer to the abolishing of the special economic zones, among which the economic, technological and exports zones, the free trade zones ect, and to the so-called "deliberate ambiguity". The first reform lead to the necessary infrastructure, the economic activity was enhanced while some of the Chinese regions managed to attract huge amounts of F.D.I. capitals, creating city regions (Zhao & Zhang, 2007:979-944). The second main reform includes the deliberate ambiguity, which change the China's attitude towards F.D.I. and made foreign capital inflow more acceptable (Smart & Hsu, 2004:544-566). These reforms were performed gradually. China became more opened to F.D.I. inflow through reducing the tariff levels and the non – tariff barriers. It has to be mentioned that nowadays most of the imports in China are duty – free (Yang, 2006:40-56).

2. China's F.D.I. and the imitation procedure

China is situated in the Asian region, where the competition for F.D.I. has become very intense during the past decade (Liu et al., 2007: 70–88). China has a positive attitude towards F.D.I. because the country uses these capitals in order to supplement the Chinese savings, but mostly because the government of the country believes that this inflow of foreign capitals will lead to external benefits and to positive spillovers for the Chinese enterprises (Qi et. al., 2009: 18-35). Besides, Chinese companies were left behind concerning the use of technological practises, compared to other enterprises worldwide. Therefore, China used imitation, skills acquisition, competition and export so as to generate positive spillovers (Gorg & Greenaway, 2004).

China has imitated other developed countries' policies so as to attract F.D.I. based on the export – oriented growth model applied by the Chinese government. Therefore, the country created export – processing zones, improved its infrastructures, applied duty exemptions policies for imported resources which were useful for the export enterprises, provided economical subsidies and tax exemption ect (Lu & Tang, 1997). Developing countries can use these four spillovers channels so as to imitate the Chinese policy and to attract more foreign capital inflows. Hence, the enterprises of the host countries can gain more knowledge on the products and the technology used from the foreign enterprises using the reverse engineering. Also, host countries can provide incentives to the skilled workers employed in Chinese multinational enterprises and to gain the technological know – how. Moreover, local enterprises can develop their R&D activity so as to develop new products and to improve the existing ones. The perceived risk of such an innovation procedure is lower for the local firms because the Chinese products and technologies have already been proven. Finally, the imitation procedure can be enhanced by the staff training, the vertical transfer of technological know – how ect (Cheung & Lin, 2004:25-44).

Another empirical study (Abeysinghe & Lu, 2003) reached to the conclusion that over the past 15 years the multiplying effect of the Chinese economy towards its neighbouring economies has been magnified. Hence, China's policies regarding F.D.I. have positive effects on the country's growth as well as on the development of the region (Abeysinghe & Lu, 2003:164-185). Also, China influences other developing Asian and regional countries because of its economy size, its openness and its rapid development. The fact that the Chinese economy grows almost three times as rapidly as other economies worldwide makes it more probable that developing countries would imitate China's attitude towards F.D.I. It is believed that the developing countries of the region do not consider China a menace because these countries also develop rapidly (Das, 2008:57-62).

Besides, as presented in table 1, the bilateral F.D.I. flows among the Asian countries have been important so far. Moreover, approximately 60% of the F.D.I. flows from East Asia have been directed to South – East Asian countries that are characterised by high average income, such as Singapore, Malaysia and Thailand (Hattari & Rajan, 2009:73-93).

	_	Average		In per co	ent to Asia
Source	Host	1997- 2000	2001- 2005	1997- 2000	2001- 2005
Hong Kong SAR	China	17,750.8	17,819.1	46.2	50.7
China	Hong Kong SAR	7,266.9	5,459.4	18.9	15.5
Singapore	China	2,706.3	2,136.7	7.0	6.1
Singapore	Hong Kong SAR	2,835.3	353.1	7.4	1.0
Singapore	Malaysia	844.1	1,133.8	2.2	3.2
Singapore	Thailand	441.7	1,381.9	1.1	3.9
Malaysia	China	290.8	316.7	0.8	0.9
Hong Kong SAR	Malaysia	272.3	296.5	0.7	0.8
Hong Kong SAR	Thailand	360.1	160.8	0.9	0.5

Table 1: Top 10 bilateral F.D.I. flows among Asian countries *

Korea	Hong Kong SAR	313.0	155.7	0.8	0.4

Source: UNCTAD in Hattari & Rajan, 2009:73-93, *amounts in \$ 100 million

However, Chinese companies were not absolutely capable of taking advantage of these spillovers channels. Hence, because of the different absorptive capability, enterprises of some of the sectors or regions gain profit from these channels. Most of them used the spillovers channels to absorb knowledge on the first production stages, at which they lagged compared to the initial production stage of firms in other countries. Though, when the difference among the production stages was great, then F.D.I. generated negative spillovers, reducing their competitiveness (Smeets, 2008: 107 - 138).

Also, spillovers effects have a positive influence on a country's economy in case there is a technological gap between foreign and local enterprises (Gerschenkron, 1962:22-45; Meyer, 2004: 259–276). The countries that will reform their attitude towards F.D.I., such as China, in order to attract foreign capitals must be able to absorb the technological knowledge deriving from China (Kokko, 1994: 279–293; Kokko et. al., 1996: 602–611). Hence, it is recommended that this gap in technology is not quite large.

Moreover, the developing countries of the Southeast Asia develop a pessimistic attitude towards the economic development of China deriving from the F.D.I. Also, neighbouring countries fear that China will attract most of the future F.D.I. inflow of the region, restraining their development (Cheong, 2001: 419– 422). Thus, the neighbouring countries have developed a competitive attitude towards China in certain categories of products because they consider that the development of China will reduce their market share both in these countries and in other countries as well (Roland-Holst & Weiss, 2005:18-35).

Besides, the economies of the Asian region are very different among each other. They are classified in different stages of development and not all of them have access to the same or to equal resources (Das, 2008: 57-62). So far, the integrated circuit (IC) industry of certain countries, such as Taiwan, has imitated the attitude of Chinese enterprises towards F.D.I. Taiwanese enterprises were offered by China land at low cost, cheap labour force, preferential tax treatment and political enforcement. Also, China offered political stability to Taiwan, bureaucratic efficiency and administrative aid to the foreign firms (Taiwan Electricity and Electronic Manufacturer Association, 2001).

Another study (Erixon & Messerlin, 2009) examines the relationship developed between the European developing countries and China. The political and trade relationships have to be re – examined and it has to paid attention on the exports as a spillovers channel because the European countries are the largest market for the Chinese economy. Therefore, the European firms should imitate the Chinese ones so as to guarantee their sales, growth and development. However, European firms have to take under consideration the consumer boycotts and the political reactions of the Chinese government that might cause problems in the imitation procedure (Erixon & Messerlin, 2009:83-85).

Apart from the European and the Asian countries, the relationship between China and the Latin American countries has been studied as well. In particular, it is proposed that the development of China and Mexico is possible to threat the U.S. development and economy. However, until the late '90s Mexico had a powerful manufacturing industry and a developed external trade, which made the Mexican government, consider China a threat, because of the country's low wages and high productivity. Nowadays, China managed to overtake Mexico as the largest trading partner of the U.S., after Canada. Therefore, China is a threat for the Mexican economy and the rest Latin American countries, which have to deal with destabilizing problems and face the China's increasing market share (Adams, 2003:36-45).

So far the U.S. has implemented some reforms so as to attract F.D.I., similar to those implemented by China in order to attract more F.D.I. Thus, U.S. government enabled the competitiveness of Arizona by improving the infrastructure and the economic and environmental conditions. Thus, the productivity of the certain state was increased, as well as the amount of F.D.I. inflow. Moreover, the state attracted small technology firms and developed collaborative relationships with several institutions (Broome, 2007: 26-32).

Finally, developing countries, especially the Asian ones, should consider the Chinese development an opportunity, not a threat. These countries are continually influenced by the development of China and it is possible that in the future they will reinforce the growth of the region. However, over the past three decades China has become very competitive and managed to become the largest recipient of F.D.I. inflow worldwide. Thus, it is possible that the competition between China and the developing countries of the region will become more intense, enhancing the spillovers effects' diffusion (Das, 2008: 57-62). Though, China has imitated the developed countries in order to attract F.D.I. and it is suggested for the rest developing countries to do the same. Besides, it is proven that the imitation of the products of the developed countries play an important role in the development of high performing countries, among which Japan and the rest recently industrialized economies of the region. Also, the imitation procedure of intermediate products produced in developed countries is the major source of productivity increase in developing countries (Kind, 2004:47-67).

Therefore, it is suggested for the rest developing countries that wish to imitate China so as to attract F.D.I. to improve their investment environment, which refers to macroeconomic and governmental factors, as well as to the infrastructure. In particular, macroeconomic factors refer to the social and political conditions of the developing country, to its inflation and interest rates, to the competition conditions ect. The governmental and institutional factors refer to the regulation, financial and taxation conditions, as well as to the legal system and the work force of the developing country ect. Finally, the last factor includes telecommunication, transportation and power infrastructure ect (Penia & Salas, 2006:70-89).

However, the multinational enterprises of the developing countries are definitely at a disadvantage because of the lack in the local knowledge and therefore it to necessary to own some ownership advantages to face this problem. Therefore by imitation the diffusion of this property of knowledge, either in technology, product or innovation process, or simply in organizing the management or administration, is considered one of the main channels through which domestic firms manage to improve productivity (Krugman & Helpman, 1994:32-45).

3. Conlusion

The past decades there is a great flow of F.D.I. worldwide which has affected the interest developing countries show towards foreign capitals inflow. There are several advantages deriving from F.D.I. and developing countries can attract foreign capital inflows via an imitation procedure. China is the largest recipient of foreign capital nowadays and the present study examines whether other developing countries can imitate China so as to attract F.D.I. and it emphasizes on the benefits deriving from F.D.I. for the Chinese economy and on the conditions under which the spillovers effects are positive.

The study reached to the conclusion that the competition in the host country, that is to say in China, between local and foreign enterprises is affected by economical factors, such as subsidies. Hence, competition can affect positively the imitation procedure in case local enterprises are motivated to used the existing resources in the most efficient way or in case the apply R & D for the technologies used. This conclusion is in consistent with the findings of Blomstrom & Kokko (1998), according to which F.D.I. enhance positive competition (Blomstrom & Kokko,1998: 247–277).

As for the skills acquisition, they can be improved via formal or informal contact of the local employees with foreign workers. Hence, the productive efficiency and the product can me improved. Also, direct observation of the technologies and the processes applies in production can be used, local firms can hire workers that were previously occupied in foreign firms, transactions with supplier of the host country can be performed ect. Hence, this study is in consistent with the findings of Motta & Ronde (2000) according to which multinational enterprises contribute to the improvement of the local work force skills.

Also, the present study reached to the conclusion that in order for the imitation procedure to be successful, two conditions must be guaranteed. The first one is the absorptive capacity of the developing countries that will imitate the China's example, which also affects the firms' operation of these countries. The second condition is sufficiency of the linkages regarding the activities between China and the developing countries' firms. Hence, these findings are in consistent with the finding of Cohen and Levinthal (1989: 569–596) and the findings of Balasubramanayam et al. (1996: 92–105), who reached to the conclusion that these two conditions are necessary for a developing country so as to attract F.D.I.

However, the study faces several limitations. Firstly, in the present study it has not been examined the degree at which spillovers effects are affected by the share of the ownership in the foreign enterprises. Hence, future studies might examine the cases of joint ventures and wholly own affiliates and the way that the imitation procedure is formed in such enterprises. Secondly, the present study has not examined how the absorptive capacity of the local enterprises affects the imitation procedure. Hence, future studies might examine the absorptive capacity, the ownership structure of the enterprises and the technological knowledge of the developing countries that is possible to imitate China so as to attract F.D.I. Finally, an interesting path for future research would therefore be to take into consideration these limitations so as to forecast F.D.I. in developing countries, as well as to examine the effect of the recent financial crisis on the imitation procedure. Moreover, future studies can examined whether there is an imitation procedure regarding the outward F.D.I. from China to other developing countries, using the level of per capital income, as proposed by Dunning (1991, 1998).

References

Abeysinghe, T. & Lu, D. (2003) China as an economic powerhouse: Implications on its neighbours. *China Economic Review*. Vol 14, pages 164-185

Adams, J. (2003) Mexico vs China, *Economic Development Journal*. Vol 2, no 4, pages 36-45

Aitken, B. & Harrison, A. (1999) Do domestic firms benefit from direct foreign investment? Evidence from Venezuela. *The American Economic Review*. Vol 89, no 3, pages 605–618

Aitken, B., Hanson, G. & Harrison, A. (1997) Spillovers, Foreign Investment and Export Behaviour. *Journal of International Economics*. Vol 43, pages 103-132

Aitken, B., Harrison, A. & Lipsey, R. (1996) Wages and Foreign Ownership: A Comparative Study of Mexico, Venezuela and the United States. *Journal of International Economics*. Vol 40, pages 345-371

Balasubramanayam, V., Salisu, M. & Sapsford, D. (1996) Foreign direct investment and economic growth in EP and IS countries. *Economic Journal*. Vol 106, pages 92– 10Bernard, A. & Jensen, B. (1999) Exceptional exporters' performance: Cause, effect or both. *Journal of International Economics*. Vol 47, pages 1–25

Blomström, M. & Kokko, A. (1998) Multinational Corporations and Spillovers, *Journal of Economic Surveys*. Vol 12, pages 247–277

Broome, B. (2007) Driving Arizona's Global Economy. *Economic Development Journal*. Vol 6, no 2, pages 26-32

Caves, R. (1974) Multinational firms, competition, and productivity in host-country markets. *Economica*. Vol 41, pages 176–193

Cevis, I. & Camurdan, B. (2009) The economical determinants of foreign direct investment (F.D.I.) in developing countries and transition economies. *e-Journal of New World Sciences Academy (NWSA)*. Vol 4, no 3, pages 110-223

Cheong Y. R. (2001) The impact of China's entrance to the W.T.O. on neighbouring East Asian economies. *China Economic Review*. Vol 11, no 4, pages 419–422

Cheung, K. Y. & Lin, P. (2004) Spillover effects of F.D.I. on innovation in China: Evidence from the provincial data. *China Economic Review*. Vol 15, pages 25–44

Cohen, W. & Levinthal, D. (1989) Innovation and learning: two faces of R&D. *Economic Journal*. Vol 99, pages 569–596

Das S. (1987) Externalities and technology transfer through Multinational Corporations: A Theoretical Analysis. *Journal of International Economics*. Vol 22, pages 171-182

Das, S. (2008) Rapid growth in China and the Asian economies: Some stylised freemarket policy inferences. *Economic Affairs*. Vol 28, no 3, pages 57-62

Dunning, J. (1991) The Eclectic Paradigm of International Production: Some Empirical Tests. *Journal of International Business Studies*. Vol 11, no 1, pages 9–31

Dunning, J. (1995) The global economy and regimes of national and supranational governance. *Business and Contemporary World*. Vol 7, no 1, pages 124–136

Dunning, J. (1998) Location and the Multinational Enterprise: A Neglected Factor. *Journal of International Business Studies*, Vol 29, no 1, pages 45–66.

Erixon, F. & Messerlin, P. (2009) Containing Sino – European Protectionism, *Economic Affairs*. Vol 29, no 1, pages 83 – 85

Fosfuri, A., Motta, M. & Ronde, T. (2001) Foreign direct investment and spillovers through workers' mobility. *Journal of International Economics*. Vol 53, pages 205–222

Franco, C. & Sasidharan, S. (2010) MNEs, technological efforts and channels of export spillover: An analysis of Indian manufacturing industries. *Economic Systems*. Vol 34, pages 270–288

Fu, X. (2008) Foreign Direct Investment, Absorptive Capacity and Regional Innovation Capabilities: Evidence from China. *Oxford Development Studies*. Vol 36, no 1, pages 89-110

Fung, K.C., Garcia – Herrero, A. & Siu, A. (2009) A Comparative Empirical Examination of Outward Foreign Direct Investment from Four Asian Economies: People's Republic of China; Japan; Republic of Korea; and Taipei, China. *Asian Development Review*. Vol 26, no 2, pages 86-101

Gerschenkron, A. (1962) *Economic Backwardness in Historical Perspective: A Book of Essays*, Belknap Press of Harvard University Press: Cambridge, MA

Glass, A. & Saggi, K (1999) Foreign Direct Investment and the Nature of R&D. *Canadian Journal of Economics*. Vol 32, no 1, pages 92-117

Glass, A. & Saggi, K. (2001) Innovation Incentives and Wage Effects of International Outsourcing', European Economic Review, 45: pages 67-86

Glass, A. & Saggi, K. (2002) Intellectual Property Rights and Foreign Direct Investment. *Journal of International Economics*. Vol 56, no 2, pages 387-410

Glass, A. & Wu, X. (2007) Intellectual Property Rights and Quality Improvement. *Journal of Development Economics*. Vol 82, no 2, pages 393-415

Globerman, S. (1979) Foreign direct investment and 'spillover' efficiency benefits in Canadian manufacturing industries. *The Canadian Journal of Economics*. Vol 12, no 1, pages 42–56

Gorg, H. & Greenway, D. (2004) Much ado about nothing? Do domestic firms really benefit from foreign direct investment'. *The World Bank Research Observer*. Vol 19, no 2, pages 171–197

Grossman, G. & Helpman, E. (1991) Endogenous product cycles. *Economic Journal*, Vol 101, no 408, pages 1214–1229

Haddad, M. & Harrison, A. (1993) Are there positive spillovers from direct foreign investment? Evidence from panel data for Morocco. Journal of Development Economics, Vol 42, no 1, pages 51–74

Hattari, R. & Rajan, R. (2009) Understanding bilateral F.D.I. flows in developing Asia. *Asian-Pacific Economic Literature*. Vol 23, no 2, pages 73-93

Hill, H. & Athukorala, P. C. (1998) Foreign Investment in East Asia: A Survey. *Asian-Pacific Economic Literature*. Vol 12, no 2, pages 23-50

Hymer, S. (1960) The international operation of national firms: a study of direct foreign investment, Ph.D. diss. Massachusetts Institute of Technology, Boston, MA

Institute of Economic Affairs (2005) Towards a Liberal Utopia?, *The Institute of Economic Affairs*, London

Ivus, O. (2011) Trade-related intellectual property rights: industry variation and technology diffusion. *Canadian Journal of Economics*. Vol 44, no 1, pages 201-226

Kind, H. (2004) Consequences of Imitation by Poor Countries on International Wage Inequalities and Global Growth. *Review of Development Economics*. Vol 8, no 1, pages 47-67

Kokko, A. (1994) Technology, market characteristics, and spillovers. *Journal of Development Economics*. Vol 43, pages 279–293

Kokko, A., Tasini, R. & Zejan, M. (1996) Local technological capability and productivity spillover from FDI in the Uruguayan manufacturing sector. *Journal of Development Studies*. Vol 32, pages 602–611

Krugman, P. & Helpman, E. (1994) Market Structures and Foreign Trade. MIT Press, Cambridge

Krugman, P. (1979) A model of innovation, technology transfer, and the world distribution of income. *Journal of Political Economy*. Vol 87, pages 253–266

Liu, L., Chow K. & Li, U. (2007) Has China crowded out foreign direct investment from its developing East Asian neighbors?. *China & World Economy*, Vol 15, no 3, pages 70–88

Lu, D. & Tang, Z. (1997) State intervention and business in China: The role of preferential policies, London: Edward Elgar

Lu, J. & Xu, D. (2006) Growth and survival of international joint ventures: an external-internal legitimacy perspective. *Journal of Management*. Vol 32, no 3, pages 427–448

Makino, S. & Delios, A. (1996) Local knowledge transfer and performance: implications for alliance formation in Asia. *Journal of International Business Studies*. Vol 27, no 5, pages 905–927

Mansfield, E., Schwartz, M. & Wagner, S. (1981) Imitation Costs and Patents: An Empirical Study. *Economic Journal*. Vol 91, no 364, pages 907-918

Meyer, K. (2004) Perspectives on multinational enterprises in emerging economies. *Journal of International Business Studies*. Vol 35, pages 259–276

Motta, M. & Ronde, T. (2000) Foreign Direct Investments and Spillovers Through Workers' Mobility, *Universidad Pompeu Fabra*, Economics Working Papers, no 258

Nanda, N. (2009) Growth Effects of F.D.I.: Is Greenfield Greener?. *Perspectives on Global Development and Technology*. Vol, 8, pages 26-47

Penia, E., & Salas, I. (2006) Investment Climate, Productivity, and Regional Development in a Developing Country. *Asian Development Review*. Vol 23, no 2, pages 70-89

Qi, J., Zheng, Y. et. al. (2009) Productivity spillovers from F.D.I. in China: Regional differences and threshold effects. *China and World Economy*. Vol 17, no 4, pages 18-35

Roland-Holst, D. & Weiss, J. (2005) People's Republic of China and its Neighbours: evidence on regional trade and investment effects. *Asian-Pacific Economic Literature*. Vol 19, no 2, pages 18-35

Romer, P. (1990) Endogenous technological change. *Journal of Political Economy*. Vol 98, pages 71–102

Segerstrom, P. Anant, T.C.A., Dinopoulos, E. (1990) A Schumpeterian model of the product life cycle. *American Economic Review*. Vol 80, pages 1077–1091

Smart, A. & Hsu, J. Y. (2004) The Chinese diaspora, foreign investment and economic development in China. *Review of International Affairs*. Vol 3, no 4, pages 544–566

Smeets, R. (2008) Collecting the pieces of the F.D.I. knowledge spillovers puzzle. *World Bank Research Observer*. Vol 23, no 2, pages 107 – 138 Taiwan Electricity and Electronic Manufacturer Association (2001) A Survey Report on the Evaluation of Investment Environment and Risks in China, *Taiwan Electrical and Electronic Manufacturers' Association (Chinese)*, Taipei

U.N.C.T.A.D. (2008) World Investment Report: Transnational Corporations and the Infrastructure Challenge. United Nations, New York and Geneva.

Wang, J.Y. & Blomstrom, M. (1992) Foreign Investment and Technology Tranfer: A Simple Model. *National Bureau of Economic Reserach Inc.* NBER Working Papers no 2958

Yan, L. (2007) Does Foreign Direct Investment Provide Desirable Development Finance? The Case of China. *China & World Economy*. Vol 15, no 2, pages 104 – 120 Yang, Y. (2006) China's Integration into the World Economy: implications for developing countries. *Asian-Pacific Economic Literature*. Vol 20, no 1, pages 40-56

Zhao, S. & Zhang, L. (2007) Foreign direct investment and the formation of global city-regions in China. *Regional Studies*. Vol 41, no 7, pages 979–994

The contribution of EU Regional Policy to structural change and to adaptation to globalization in less developed regions

Alexandros Karvounis1

¹SEERC/Department of Geography, University of Sheffield, akarvounis@seerc.og

Abstract. This paper presents work in progress in relation to a research programme aimed at assessing interventions financed in the framework of Objective 1 programmes in less developed regions, which contributed to supporting regional structural change and to enabling adaptation to globalisation over the 2000 – 2006 period. The purpose is to compare through a systematic analysis the contribution of Structural Funds' support to manufacturing firms' performance in similar economic geography profile regions but with different technological level. For this reason, one region from UK (South Yorkshire) and one from Greece (Kentriki Makedonia) and their respective Objective 1 programmes are examined thoroughly. The theoretical and methodological framework of this research is analyzed as well as the immediate priorities, next steps and some early findings.

Keywords: Regional Policy, structural change, globalization, Objective 1, Kentriki Makedonia, South Yorkshire

1 Introduction

This paper presents work in progress in relation to a research programme aimed at assessing interventions financed in the framework of Objective 1 programmes in less developed regions, which contributed to supporting regional structural change and to enabling adaptation to globalisation over the 2000 – 2006 period. The purpose is to compare through a systematic analysis the contribution of Structural Funds' support to manufacturing firms' performance in similar economic geography profile regions but with different technological level. For this reason, one region from UK (South Yorkshire) and one from Greece (Kentriki Makedonia) and their respective Objective

1 programmes are examined thoroughly. On the basis of the analysis, the research will draw up recommendations for future investment by the Structural Funds expenditure in support of manufacturing firms in the assisted regions, exploring new forms of policy improvement.

The paper starts with a basic context, reviewing the literature and identifying the most important key issues and gaps. The theoretical and methodological framework of this research is presented. The paper then provides more information on the ways in which data have been collected and processed too date (both qualitative data from interviews and quantitative secondary data) and outlines the immediate priorities and next steps.

2 Literature review

The aim of this section is to give an overview of the literature on regional development theories, structural change and globalisation. The research programme aims to add to previous work analysing the effects of structural funds in regional performance in a globalised context.

2.1 Structural change at regional level

The economic discussion upon structural change has been sustained at both national and regional level. An extensive number of academic work has discoursed on the degree and the plausible effects of specialization and structural change at the national level in the area of Europe (Midelfart-Knarvik, Overman 2002, Greenaway and Hine 1991, Krugman 1991, Helg et al. 1995, Peneder 2003, Fagerberg 20008). The regional dimension has been largely addressed in Ezurra, Pascual Rapun (2006), Molle (1996), Hallet (2002), Alves et al. (2007), Robson (2006), Le Blanc (2000) etc. The latest trends in structural change analysis have been further encouraged and strengthened by the debate on technological change and its impact on growth, as well as stimulated by new evolutions in mainstream economic literature (e.g. Lucas, Romer, Grossman and Helpman 1991) and the rise of the New Economy. Evolutionary economics put a considerable emphasis on technological diffusion, adaptation, economic interaction at the firm level and the role that the institutional, historical and cultural environment plays in certain locations (Freeman and Perez 1988, Nelson and Wright 1992).

The conceptual premises of structural change analysis have consequently been restrengthened by the dimension of space and the integration of geographical factors as relevant notions of national innovation systems (Nelson 1993, Lundvall 1992) or regional innovation systems (Cooke and Morgan 1998; Cooke 2001; Carlsson 2004; Edquist 2004; Iammarino 2005; Boschma and Frenken 2005; etc.).

The major part of the latest discussion on regional innovation and growth inequalities cites the 'Regional Innovation Systems' approach (RIS). The 'innovative milieu' that avails from knowledge spillover effects between regional actors takes into consideration the interconnections of the socio-economic frame of reference and the interplay of economic actors as key agents of regional innovation capacity (Cooke 2001, referring to Maillat and Creviosier). Irrespective of the rich amount of literature

dedicated to determine and estimate the variety of dimensions of this notion and also to attempt to deviate from the conventional notions of innovation and economic performance, and albeit the concept has been supportive to policy support at both the regional and the EU level, the innovation systems approach is still vague.

2.2 Globalization effects

Objective 1 regions and regions in the periphery are notably vulnerable to globalisation (EC 2009). The negative consequences of globalisation and relocations have been widely examined at the national (Rojec et al. 2008, OECD 2006) as well as the sectoral level (OECD 2005; Bradford et al. 2005). Even though it is easy to transfer and apply the outcomes of studies at the national level to the Objective 1 regions, sectoral studies could make an important contribution to the better understanding of context and challenges at the regional level.

2.3 Review of the most approved theories of regional development to be applied to evaluate Objective 1 programme effects

In the periods 1994-1999 and 2000-2006 Structural Funds in Objective 1 regions were actually appropriated to deal with structural problems. Academic developments beyond the bounds of mainstream economic theories, such as those of Leontief (1953), Goodwin (1976, 1983) or Pasinetti (1985), coupled with extensive discussions regarding old industrial areas and the decay of traditional manufacturing (Steiner et al. 1983, Carney et al. 1980),40 have all contributed to reinforce discussion of structural change at the regional level since the mid-1980s and offered cogent arguments for Structural Funds interventions beyond Objective 1 regions.

The initial years of Structural Funds were joined with vivid discussion of endogenous regional development. Dansons (2003) brings a stimulating analysis of mostly compatible models of regional development and innovation that all influenced policy discussion to various degrees. Some of the approaches and theories introduced include the "regional innovation systems" approach, the "innovative-milieu" approach, the "industrial district" approach, the "local production systems" approach, the "learning region" approach, and the regional "network theory".

2.4 Regional assistance and firm performance

Taking into consideration that the grant of capital subsidies is a critical tool of regional and industrial policy (Armstrong, 2001; Armstrong and Taylor, 2000; Holden and Swales, 1996; Swales, 1981; Wren, 1994), it is astonishing that there are not so many studies dealing with their impact on firm performance. At sector level, Beason and Weinstein (1996) extensively analysed the utility of varied industrial policy instruments in Japan, including capital subsidies, and against traditional wisdom, they discovered that a disproportionate amount of the targeted Japanese took place in low growth sectors and sectors with reducing returns to scale. Nonetheless, the discussion over assessment of investment subsidies on firm performance has been

confined in the literature. Tzelepis and Skuras (2004 and 2006) have examined the impact of investment subsidies on firm performance.

3 Data and Methodology

As shown in the context section, there are areas of research to be built on and gaps that could be addressed. The main aim of this research is to appraise whether structural fund interventions have assist manufacturing firms in less developed regions to adapt to globalization pressures and to support structural change.

The analysis contains a chain of steps (see Fig. 1) distinguished into stages as well as addressing aim and objectives in turn. How this work works in practice will be explained below in greater detail. So far, econometric studies have dominated the literature on effects of Structural Funds on regional performance (EC 2005, EKOS Consulting 2004, Martin R. & Tyler P. 2006); however, this method prefers to ignore how the process takes place and fails to answer the other popular question of "why". Therefore, in order to eliminate this defect, this research employs as a supplementary semi approach (see Fig. 1) an explanatory survey using structured interviews/interviewer administered questionnaires. This is preferable as approach, as it provides detailed and accurate information about the non qualitative research questions that need to be answered. With the input of the survey study, this research can be a valuable contribution to the ongoing literature and can fill in the gap of the econometric studies.

In conclusion, the research uses a combination of quantitative and qualitative methods, including

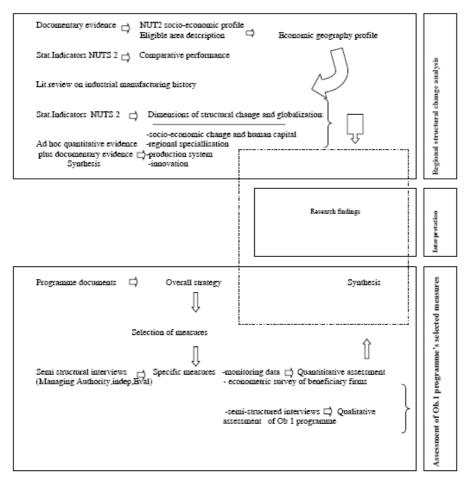
a) Extensive literature review to identify the theories behind Objective 1 Programmes

b) Comparative analysis to assess regional performance, structural change and globalisation pressures in the two case study regions, their national contexts and their respective Objective 1 Programmes

c) Econometric analysis using one way fixed effect model with a firm panel data (Baltagi 2009) to examine the effect of structural assistance on manufacturing firms' performance.

d) Interview survey with semi structured questionnaires with the key people involved in the delivery and implementation of the programmes.

Figure 1: Methodological framework



The suggested methods were selected on the basis of a scrupulous and critical review of the literature:

a) Literature review

Official policy documents literature review manifested that is not simple to distinguish which theories lie behind Objective 1 Programmes. The use of NVivo Software textual analysis in these policy documents will reveal the relation between the terms (e.g five drivers of growth - physical capital skills, innovation, competition and enterprise) of theories which are not clear in Obj. 1 policy documents. This approach will add some innovative features to the ongoing literature, which is based on conventional qualitative methodologies.

b) Comparative analysis

The analysis is premised on comparative statistical data accesible at NUTS 2 level from harmonised sources, mainly EUROSTAT and to a certain extent to National Statistic Services. The analysis will evaluate regional performance against national performance in the 2000-2006 period. This is valuable in order to assess in what extent the economic trends detected are region-specific (CSIL 2008). Regional performance will additionally be tested (Salter Graphs- Monfort 2008) against the EU average (comparison between regional, national and EU-15 average performances considering the following: GDP per capita; real growth rate of GDP per capita; GVA; real growth rate of GVA; rate of employment; rate of unemployment):Moreover, it will emphasize on particular parts of the programmes reviewed, for instance, measures co-financed by the ERDF that appear particularly compelling for measuring their effects in reinforcing manufacturing change and facilitating adaptation to globalisation.

c) Regression modelling

A regression approach is adopted based on the well established one-way, fixed effects, unbalanced panel regression model (Greene, 1997). The panel data is assembled by both the manufacturing firms that obtained EU assistance during the 2000-2006 period and all the manufacturing firms that perform in the two regions (RQ 4). In this case, independent control variables are size, export intensity and ERDF financial assistance (dummy variable) of manufacturing firms and the dependent variable are firm's performance (profitability and growth). The fixed effects panel data model is better cohered with the suggested data set, which is not limited to a small sample of the firms in the sector, but rather includes pretty a large number of them (Tzelepis, and Skuras 2004, Balsari and Ucdogruk, 2008).

d) Interview survey

Interview survey with the key people involved in the delivery and implementation of the programmes has been almost completed both in South Yorkshire and Kentriki Makedonia. These people were identified and selected based on the role they played during the implementation of the programmes. Their contribution is highly important for the qualitative assessment of the relative Objective1 Programmes.

Data for this work came from three major databases (see Table 1). Firstly cross sectional data were collected from EUROSTAT and the National Statistic Services of UK and Greece in order to assess regional performance, structural change and globalisation pressures in the two case study regions and their national contexts (comparative analysis).

Table 1: Datasets

Dataset	Year	Description
UK Census of	1991,2001	Global coverage of the UK
Population, ONS		population on a wide range
		of relevant variables

Greek Census of	1991,2001	Global coverage of the
Population, NSSG		Greece population on a
		wide range of relevant
		variables
Eurostat	2002 - 2006	Regional Statistics
Kentriki Makedonia	2000-2008	List of assisted
Operational Authority		manufacturing firms
Monitoring data		
Yorkshire and Humber	2000-2008	List of assisted
Government Office		manufacturing firms
Monitoring data		
Amadeus Database (for	2002 -2010	Individual firms
Greek firms)		characteristics from all
		manufacturing firms in
		Kentriki Makedonia
Fame Database (for UK	2002 -2010	Individual firms
firms)(Alternative to		characteristics from all
AMADEUS)		manufacturing firms in
		South Yorkshire

Source: Own elaboration

Secondly Data on individual firm characteristics for the period of 2002-2010 were derived from the global business database AMADEUS maintained by the private financial and business information Service Company called Bureau Van Dijk. The annual AMADEUS directories provide key elements from the published balance-sheets of almost all plc and Ltd firms operating in all sectors of economic activity in UK and Greece. The third database is being elaborated from the list of beneficiary manufacturing firms that have received structural assistance in South Yorkshire and Kentriki Makedonia. The above data are used to develop a dataset with assisted and non assisted manufacturing firms and their financial performance (Econometric analysis).

4 Empirical analysis

a) Comparative analysis

The comparative analysis to assess regional performance, structural change and globalisation pressures in the two case study regions has shown so far that the economic performance of the regions has improved during the 2000- 2006 period (table 9). Concerning Yorkshire & the Humber, the region still remains below the UK average for per capita GDP. The weakest sub-regional economy, that of South Yorkshire, has shown most improvement, due in no small part to the investments made during the Objective 1 Programme. The percentage on GDP change during 2002-2006 rises up to 15%, where the financial assistance of Objecive 1 reached 11% of the 2001 regional GDP (table 8). South Yorkshire needs to continue restructuring

and attracting investment to replace lost jobs and meet the employment needs of a growing economy (GoYH, 2007)

regions						
Regional gross domestic product (million EUR), by NUTS 2 regions						
						CHANGE
	2002	2003	2004	2005	2006	2002-2006
Kentriki Makedonia	21970	23882	25837	27396	29550	35%
South Yorkshire	27683	26589	29026	30228	31807	15%

 Table 2 - Regional gross domestic product (million EUR), by NUTS 2

 regions

Source: Eurostat, Own elaboration

Kentriki Makedonia seems to perform better during this period of time although the financial assistance was lower in relation to GDP. In more detail, Kentriki Makedonia increased its regional GDP around 35% from 2002 till 2006, although Objective 1 assistance amounted around 7% of the 2001 regional GDP. In absolute numbers, Objective 1 assistance was around double in South Yorkshire compared to Kentriki Makedonia, but the last one seems to reduce the gap in regional GDP with South Yorkshire in 2006.

 Table 3 - Financial allocation objective 1 - percentage of GPD (million euros for year 2001)

	FINANCIAL ALLOCATION OBJECTIVE 1	PERCENTAGE OF GDP (million euros for year 2001)
Kentriki		
Makedonia	1547,198	7%
South		
Yorkshire	3088,2307	11%

Source: Single Programming Documents, Eurostat, Own elaboration

b) Data collection and econometric model

Data collection has almost completed and most of the data concerning financial performance of manufacturing firms in the two case studies were downloaded from FAME (Amadeus) International database. For 1342 manufacturing companies in Kentriki Makedonia and 344 manufacturing companies in South Yorkshire, respective datasets have been developed for each case study, including the same variables that can descript the financial performance of the companies (e.g return on assets, annual turnover, profit margin, number of employees)

Serious problems have risen in the South Yorkshire case study where there is no list of beneficiary firms from the Yorkshire and Humber Government Office Monitoring data. This is caused by the fact that the structure of administration of the Objective 1 Programme was different in South Yorkshire and the actual delivery of assistance was a responsibility of intermediate bodies like Bussiness Link or South Yorkshire Investment Fund and not by the Yorkshire and Humber Government Office. In addition, intermediate bodies are refused to provide data for their beneficiary firms due to consumer rights protection. This problem has been addressed so far by collecting newsletters and brochures of the intermediate bodies which highlight their success stories of Objective 1 supported firms.

c) Interviews

Interviews took place in phases (from December 2009 since now) both in South Yorkshire and Kentriki Makedonia. They are based on a semi structured questionnaire and addressed to key people involved in the delivery of Objective 1 programmes in the two case studies. The sample was selected and based on the snowball method. The main key personnel from the two Objective 1 Directorates proposed other possible experts that were involved directly or indirectly with the Objective 1 Programmes. These were people from private sector and industry, academics and government officials. Ten people from South Yorkshire and seven from Kentriki Makedonia have been interviewed so far. The interview process will be completed at the end of the year and qualitative analysis of the questionnaires will commence immediately using the software NVivo. The last is a valuable instrument for detecting symbolic or metaphorical relations between spans of text or for revealing trends within the data that might not compulsorily be discovered by statistical text analysis methods.

d) Immediate priorities and next steps

The completion of the panel dataset with the assisted and non-assisted manufacturing firms in the two case studies is the prerequisite for running a fixed effect econometric model to examine the effect of structural assistance on manufacturing firms' performance. The statistical software EVIEWS will be used for the quantitative statistical analysis.

The next step includes the qualitative analysis of interview survey in order to generalise findings from the two case studies and draw up recommendations for future investment by the Structural Funds expenditure in support of manufacturing firms, exploring new forms of policy improvement.

5 Conclusions

Comparing development strategies in two regions like South Yorkshire and Kentriki Makedonia, which differ in geographical, economic and social terms, makes it difficult to assess to what extent any economic success is the result of the implementation of the policy or of the economic dynamic of the region (Rodríguez-Pose, A. 2000). Current research on programme implementation (EEO Group 2002, Kantor-Eurotec 2008), though still in an early stage, indicates that many of the deficiencies of previous programmes have not been the result of ignorance or conservatism on the part of the planners. According to Plaskovitis (2006), they rather reflect what is actually demanded on the ground from local populations and

programme beneficiaries. Wherever we meet a more ambitious and innovative measure there are huge difficulties in finding users to absorb the allocated funds. In this sense it is probably better to prepare a strategy for changing the wider development culture and stereotypes, rather than pressing for a top down programme modernization.

So is it possible to generalise the experience of these regions in other EU regions? For example, there are tangible effects from Objective 1 funding towards job creation, but there are many doubts about whether the funding has been adequate for the challenges it has faced (Taylor and Wren 1997:845) or is being used optimally (Gripaios and Bishop 2006:950). Some early research findings show that the impact and effectiveness of the strategies is largely determined by progress in the following key respects: 1) structural reforms in the labour, goods and services markets (G.S.I.D. 2005), 2) continuity, 3) mobilisation of the private sector in all regions, 4) administrative capacity, 5) scale of projects and 6) targeting.

References

Abraham, F. and van Rompuy, P. (1995). Regional Convergence in the European Monetary Union, Papers in *Regional Science*, Vol. 74, No. 2, April 1995.

Armstrong H. W. (1995), Convergence among Regions of the European Union, 1950-1990, Papers in *Regional Science*, Vol. 74, No. 2, April 1995.

Armstrong H. W. (1995). Convergence among Regions of the European Union, 1950-1990, Papers in *Regional Science*, Vol. 74, No. 2, April 1995. Armstrong, H. and Taylor, J. (2000), Regional economics and policy, 3rd edition, Blackwell.

Barro, R.J. and Sala-i-Martin, X. (1992). Convergence, *Journal of Political Economy* 100, pp.

Barro, R.J. and Sala-i-Martin, X. (1991). Convergence Across States and Regions, Brookings Papers on *Economic Activity* I, 107-182.

Barro, Robert, and Xavier Sala -i-Martin (1991), Convergence across states and regions, Brookings papers on *Economic Activity*, no. 1 / 1991, pp. 107-182.

Blomström M. (2006) Study on FDI and regional development, Copenhagen, *Economics*, vol. 70, no. 1, pp. 65-94.

Borts, G. (1960). The equalization of returns and regional economic growth, *American Economic Review*, June.

Borts, George H., and Jerome L. Stein (1964), Economic growth in a free market, Columbia:

Bradford J. J., and Kletzer L. (2005), Tradable Services: Understanding the Scope and Impact of Service Offshoring, Brookings Trade Forum 2005

Carlsson, B. (2004) Innovation Systems: A Survey of the Literature from a Schumpeterian Perspective, Paper presented at the International J.A. Schumpeter Society Conference, Milan, Italy.

Carney, J., Hudson, R. and Lewis, J. (eds. 1980) Regions in Crisis, London: Croom Helm. Columbia University Press.

Cooke P. (2001): Regional innovation systems, clusters and the knowledge economy, *Industrial and Corporate Change*, Vol.10, 4, pp. 945-974

Cooke P., Morgan K. (1998) The Associational Economy: Firms, Regions and Innovation, Oxford; Copenhagen Economics.

Egger, H., Egger P. (2000) Outsourcing and skill specific employment in a small economy: Austria and the fall of the iron curtain, Universität Innsbruck - Institut für Volkswirtschaftslehre, *working paper no. 0024*

EKOS Consulting (2004), The impact and added value of EU structural funds in Yorkshire and the Humber, Leeds: EKOS Consulting

European Commision (2005) The impact of cohesion Policy in Member States, Evaluation Unit, Luxembourg.

European Commission (2007), Fourth report on economic and social cohesion, European Commission, Luxembourg.

European Commission (2008), Fifth progress report on economic and social cohesion, European Commission, Luxembourg.

European Commision (2009), Regions 2020: Globalization challenges for European Regions, Background document to Commission staff working document sec(2008)2868 final *Regions 2020 : an assessment of future challenges for EU regions, European Commission*, Luxembourg.

European Enterprise Organisation - EEO GROUP (2002), Ex post evaluation of CSF in Objective 1 regions, Country report, Greece

Ezcurra R. Pascual P., Manuel Rapun, (2006) Regional Specialization in the European Union, *Regional Studies*, Vol. 40.6, pp 601-616

Fagerberg J (2000). Technological Progress, Structural Change and Productivity Growth: A Comparative Study. *Structural Change And Economic Dynamics*, 11(4): 393-411

Freeman C., Perez C. (1988) Structural Crisis of adjustment, business cycles and investment behaviour, in: Dosi G. Freeman C., Nelson R., Silverberg G., Soete L. (eds. 1988) *Technical Change and Economic Theory*, Pinter London

Greenway, D. Hine R. (1991). Intra-industry specialisation, trade expansion and adjustment in the European economic space. *Journal of Common Market Studies* 29: 603-622

Hallet, M. (2002). Regional Specialisation and Concentration in the EU. In J. R. Cuadrado and M.

Helg R., Manasse P., Monacelli P., Rovelli R. (1995) How much (a)symmetry in Europe ? Evidence from industrial sectors, *European Economic Review 39*, 1017-1041

Hirschman, A. (1958) *The Strategy of Economic Development*, New Haven: Yale University Press.

Kaldor, Nicolas (1970), The case for regional policies, *Scottish Journal Political Economy*, vol. 17, pp. 337-348.

Klein, J. (2001) A Critique of Competitive Advantage, Critical Management Studies Conference, Manchester

Krugman, P. (1991). Geography and Trade. Cambridge: MIT Press.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy* 99: 483-499.

Krugman, P. (1991a) Increasing returns and economic geography, Journal of Political Economy, 99:484-499

Krugman, P. (1991b) Geography and Trade, Cambridge: MIT Press

Krugman, P. (1993) Space: The Final Frontier, *Journal of Economic Perspectives*, Vol. 12, No. 2, pp. 161-174.

Krugman, P., Venables, A. J., (1990) Integration and the Competitiveness of Peripheral Industry, in: *CEPR Discussion Papers 363*, C.E.P.R. Discussion Papers

Krugman, Paul and Anthony Venables (1995), Globalization and inequality of nations, *Quarterly Journal of Economics*, vol. 110, pp. 857-80.

Le Blanc G. (2000) Regional Specialization, Local Externalities and Clustering in Information Technology Industries, CERNA, Centre d'économie industrielle Ecole Nationale Supérieure des Mines de Paris

Marshall, Alfred (1920), Principles of Economics, London: MacMillan.Martin, R. & Tyler, P., (2006), Evaluating the impact of the structural funds on objective 1 regions: An exploratory discussion, *Regional Studies*, Taylor and Francis Journals, vol. 40(2), pages 201-210, April.

Molle, W. (1996) The regional economic structure of the European Union: an analysis of long term developments, in K. Peschel (ed.): *Regional Growth and Regional Policy Within the Framework of European Integration*. Heidelberg: Physica-Verlag.

Molle, W. and Boeckhout, S. (1995). Economic Disparity under Conditions of Integration - A Long Term View of the European Case, Papers in *Regional Science*, Vol. 74, No. 2, April 1995.

Myrdal, G., 1957, Economic Theory and Underdeveloped Regions, London: Duckworth&Co.

Nelson R. (ed. 1993) National Innovation Systems: A Comparative Analysis, Oxford University Press, New York

North, D.C. (1955), Location and regional economic growth, *Journal of Political Economy*, vol. 63, pp. 243-58.

OECD, National accounts, Paris: OECD.

OECD (2005) Trade and structural adjustment policies in selected developing countries, *Working Papers* No. 245, Paris.

OECD (2005a) Building Competitive Regions. Strategies and Governance, Paris.

OECD (2005b) The Governance of Innovation Systems: volume 1 : Synthesis report, Paris 2005

OECD (2005c) The impact of structural policies on trade-related adjustment and the shift to services. Economics Department *Working Papers* No. 427, Paris.

OECD (2005d) Trade and structural Adjustment – Embracing Globalization, Paris 2005

OECD (2006a) The OECD Input-Output database: 2006 Edition, Paris.

OECD (2006b) Working Party on the Information Economy - The share of employment potentially affected by offshoring – an empirical investigation; DSTI/ICCP/IE(2005)8/FINAL, Paris

OECD (2006c) Working Party on the Information Economy - The share of employment potentially affected by offshoring - an empirical investigation; Directorate for Science, Technology and Industry; Paris.

Peneder M (2003) Industrial structure and aggregate growth, *Structural Change and Economic Dynamics* (2003) 14, (4), 427-448 113

Perroux, François (1955), Note sur la notion de pôle de croissance, *Economique appliquée*, vol. 1-2, pp. 307-320.

Petrakos G. and Saratsis Y. (2000), Regional inequality in Greece, Papers in *Regional Science*, Vol. 76: 57-74

Raines P., (2001), Local or National competitive advantage? The Tensions in Cluster Development Policy), *Regional and Industrial Policy Research Paper*, EPRC, University of trathclyde, Glaskow.

Reckon LLP (2006) A sectoral survey of relocation: a factual background; client: European Economic and Social Committee (EESC); Consultative Commission on industrial change (CCMI); London

Robson M. (2006) Sectoral Shifts, Employment Specialization and the Efficiency of Matching: An Analysis

Rojeca M., Damijanc J.P (2008) Relocation via foreign direct investment from old to new EU member states: Scale and structural dimension of the process, *Journal of Structural Change and economics*, 4/ 2008

Solow (1956) A Contribution to the Theory of Economic Growth, *Quarterly Journal Economics*, Vol. 70, p.65-94.

Tichy, G. (1997) Are today's clusters problem areas of tomorrow? in: Steiner M. (ed. 1997), *Competence Clusters. Workshop Report*, Graz: Leykam, pp. 94-100

Tiebout, Charles M. (1956), Exports and regional growth, *Journal of Political Economy*, vol. 64, pp. 162.

Tondl, G. (2001), Convergence after Divergence? Regional Growth in Europe, Springer Verlag: Wien / Heidelberg / New York, 350 pages, 2001

Conspicuous Consumption: Essential Concepts

Armando Memushi*

*Teaching Assistant at the Department of Economics, Faculty of Economy, University of Tirana (E-mail: armandomemushi@yahoo.com)

Abstract: Since Veblen introduced the concept of conspicuous consumption in 1899, this term is often used in a vague descriptive sense to refer to any non utilitarian forms of consumption, or simply to that which is judged extravagant, luxurious, or wasteful. Developing countries may have relatively higher levels of conspicuous consumption, because of extreme income inequalities. By contrast, there is a quasi total lack of research in this field in Albania. Furthermore, in spite of the massive volume of "foreign" research, the exact nature of conspicuous consumption has not been fully investigated. Drawing upon existing relevant literature, this paper briefly discusses several aspects of the phenomenon. Changes in conspicuous consumption (and its definition) during time are presented. Some of the reasons behind conspicuous consumption and motivations that lead consumers toward it are explained, drawing upon relevant analyses made by researchers coming from economics, marketing, sociology and psychology. Identifying which factors may have an impact on the level of conspicuous consumption of a household, is maybe the main and the most difficult of the problems discussed below.

This study aims to be only the starting point of future more extended examinations in upcoming research. A questionnaire will be soon conducted to find out the main reasons and motivations lying beneath conspicuous consumption of luxury goods in the Albanian environment. Moreover, an econometric analysis of the level of such consumption in the Albanian households will be certainly developed, using data from the Living Standards Measurement Surveys, for more than 3500 households.

Key Words: *Conspicuous Consumption, History, Motivations, Factors, Signaling.*

JEL Classification: D03, D11, Z13.

1 Introduction

To be, or to be seen, that is the question ...

The term *conspicuous consumption* is familiar to most economists, marketers, sociologists and psychologists and it has also become part of everyday language. This term is often used in a vague descriptive sense to refer to any non utilitarian forms of consumption, or simply to that which is judged extravagant, luxurious, or wasteful.¹ Even though, the number of scientific studies conducted in this field is still relatively low, maybe due to the interdisciplinary character of the topic. The study of such a phenomenon may be a subject of common interest, especially in a developing country like Albania. Developing countries may have a relatively higher level of conspicuous consumption, because of extreme inequalities in income. The relatively frequent appearance of such a behavior in the Balkans in general, seems to confirm this idea. Thus, it sounds particularly fascinating to investigate reasons, factors and other aspects concerning this behavior. This may help to better understand the coexistence of the impossibility to fulfill sometimes even the basic needs and the insistence to buy extremely expensive products simply to show off, on the other side.

The concept studied here has its origins in the work of Thorstein Bunde Veblen (1857–1929), a Norwegian-American economist and sociologist, one of the founders of institutional economics. He wrote his The Theory of the Leisure Class more than one hundred years ago (1899), based entirely on observation. Even though, it still represents a powerful critique of the neoclassical theory of consumption. It contrasts the neoclassical approach, which assumes the individual's maximization of utility according to exogenous preferences, by developing an evolutionary framework in which there are also social factors. Thus, preferences are determined socially depending on the positions of individuals in the social hierarchy.² Veblen is one of the first theorists who argued that consumption is actually a process of socialization and goods have a function as markers of social class.³ He argues that "wealthy individuals often consume highly conspicuous goods and services in order to advertise their wealth, thereby achieving greater social status".⁴ He named this kind of unnecessary and unproductive expenditures, as conspicuous consumption. Hence, he ignores that even those from the lowest scales of the hierarchy can also engage in conspicuous consumption. Therefore, conspicuous consumption is seen mistakenly as an exclusive instrument for the rich. Finally, even after a closed reading and analysis of the different arguments given in The Theory of the Leisure Class, the formulation of Veblen's most famous concept is not sufficiently clear to permit any general agreement on its definition.⁵

¹ Campbell, C. (1995)

² Trigg, A. B. (2001)

³ Patsiaouras, G. (2010)

⁴ Bagwell, L. S. & Bernheim, B. D. (1996)

⁵ Campbell, C. (1995)

Reformulating arguments given by Veblen, conspicuous consumption can be defined as the consumption of goods and services that is motivated mainly by secondary (conspicuous) utility, i.e. utility deriving by evidence or social confirmation of the consumer's relative ability to pay.⁶ Consumers make their decisions to buy and consume "conspicuous" goods, paying attention not only to the material needs that these products can satisfy, but also to social needs such as prestige.⁷ Hence, individuals (not only the rich ones, but also middle class and working class members) involve in a competitive race for gaining status or prestige. Examining modern relevant literature, Scheetz defines conspicuous consumption to be any consumption which purpose is that of showing off wealth to others when the good is publicly consumed, "ranging from applying an expensive lipstick in public to driving an expensive car".⁸ So, the concept created by Veblen, is generally seen by researchers (mainly economists) as a process or means to achieve or maintain status and other social benefits, mainly through the consumption of luxury goods. It is still almost impossible to give an exhaustive definition of it, maybe due to the mix nature of the phenomenon, which brings the need of multidisciplinary analysis combining viewpoints from economics, marketing, sociology, psychology and even biology. "Consumer behavior is often too complex to be handled by economics alone and, if done, may severely limit the scope of findings."9

Drawing upon existing relevant literature, this paper briefly discusses several aspects of the phenomenon. Changes in conspicuous consumption (and its definition) during time are presented. Some of the reasons behind conspicuous consumption and motivations that lead consumers toward it are explained, drawing upon relevant analyses made by researchers coming from economics, marketing, sociology and psychology. Identifying which factors may have an impact on the level of conspicuous consumptions of a household, is maybe the main and the most difficult of the problems discussed below. This study aims to be only the starting point of future more extended examinations in upcoming research. A questionnaire will be soon conducted to find out the main reasons and motivations lying beneath conspicuous consumption of luxury goods in the Albanian environment. Moreover, an econometric analysis of the level of such consumption in the Albanian households will be certainly developed, using data from the Living Standards Measurement Surveys, for more than 3500 households.

⁶ Basmann, R. L. et al. (1988)

⁷ Amaldoss, W. & Jain, S. (2005)

⁸ Scheetz, T. (2004)

⁹ Chaudhuri, H. R. & Majumdar S. (2006)

2 Literature Review

2.1 History of Conspicuous Consumption

"The wish to become proper objects of this respect, to deserve and obtain this credit and rank among our equals, may be the strongest of all our desires." Adam Smith

Veblen was not the first...Few of the concepts or theories of Thorstein Veblen figure in ongoing debates or are employed in contemporary research; the one obvious exception is his concept of conspicuous consumption.¹⁰ Even though he is the creator of the concept, it is certainly not the first economist to write about such a behavior. Adam Smith argued that "it is not wealth that men desire, but the consideration and good opinion that wait upon riches". Karl Marx already noted that ostentatious consumption can be seen as a business necessity and a signal of trust, to attract further capital.¹¹

But, it must be said that conspicuous consumption was not considered as an interesting area of study, for a long time. Even Veblen himself was first not considered as a "real economist" and his ideas related to the consumer theory were rejected as non scientific. Only in the XX century, mainly after the Great Depression, economists and researchers from other fields began to make efforts to explain such a phenomenon and its implications on the economy.

Seeing a woman all dressed up for a trip to the city, Socrates remarked: "I suspect that your trip is not to see the city, but for the city to see you..." Socrates (470-339 B.C.)

A long story of...conspicuous consumption. Conspicuous consumption is not a recent phenomenon. The early civilizations were plentiful of material displays of power. The upper classes "decorated" themselves with jewelry that was fashioned from rare and exotic materials.¹² Rulers sponsored elaborate ceremonies, which often required special equipment and in which many people were employed as actors and participants. During them, large quantities of food and manufactured goods were destroyed as offerings or distributed among the participants. The fact that conspicuous consumption was satirized by Petronius, and mocked in several plays by Plautus demonstrates that these issues were not unknown to wider audiences of that time.¹³ Attitudes and even laws concerning conspicuous consumption and luxury have been presented by Ray, demonstrating the various desires exhibited by some sectors of Roman society. He also makes comparisons, giving examples from the Ancient Greece world. Signs of similar wasteful consumption can be found even among ancient Polynesians.¹⁴ In the ancient civilizations large amounts of resources were

¹⁰ Campbell, C. (1995)

¹¹ Marx, K. (1867)

¹² Murra, J. V. (1962)

¹³ Ray, N. M. (2009)

¹⁴ Leach, H. (2003)

also invested in elaborating and maintaining complex systems of writing, calendric systems, astrology, divination and other religious traditions, being that a large part of the "great tradition".¹⁵ This served to emphasize the cultural gap between the upper and lower classes, the social and political disadvantage of the subordinate ones. During the early phases of Egyptian civilization, for example, even royal effigies were some times placed where they could be seen by ordinary people, aiming through these means to give the message of differentiation and exclusion.

Monumental architecture in early civilizations is also a typical example of conspicuous consumption. Power was symbolized and reinforced by the large scale on which processional ways, palaces, throne rooms, temple platforms, and royal tombs were constructed.¹⁶ Signs of such "wasteful" consumption can be easily found in the Aztec culture and the behavior of Aztec rulers. For example, pleasure parks and palaces and their role in courtly life can be described and interpreted as pure examples of conspicuous consumption and vehicles for status rivalry between the two great noble lineages of Aztec Mexico, rulers of the two major cities, Tenochtitlan and Texcoco.¹⁷ These are only a few examples showing the presence of conspicuous consumption and related phenomena in early civilizations all over the world. As it can be seen, these behaviors were strictly related to rulers and other members of the elite classes (nobility).

During the XVI-XVII centuries, in Europe production was considered to be the main economic mechanism and impulse behind the advancement of a wealthy nation and luxury consumption was therefore condemned.¹⁸ However, a new reality was created because of developments made in communication, transportation and construction after the Industrial Revolution. In this new environment, increasing income, new employments and geographical mobility generated an ambitious middle class, whose members participated with enthusiasm in the game of ownership and ostentatious behavior, a privilege of the aristocratic elites until then.¹⁹ This situation inspired Veblen later on to conduct research and to write his "theory", aiming to explain and mainly criticize the behavior of the *nouveaux riches*.

In the modern societies, conspicuous consumption continued to be present, with periods of rising and falling down, related to social, economic and political background.²⁰ It is still present and even more discussed nowadays. Living in a "society of image", consumers have even more incentives to conspicuously consume. Furthermore, this phenomenon is no more a privilege of the upper classes, but a common behavior even among the "masses".²¹ Maybe, one of the factors staying behind of the recent financial and economic global crisis can be found in behaviors

¹⁵ Trigger, B. G. (1990)

¹⁶ Trigger, B. G. (1990)

¹⁷ Evans, S. T. (2000)

¹⁸ Patsiaouras, G. (2010)

¹⁹ Page, C. (1992)

²⁰ Patsiaouras, G. (2010)

²¹ Chaudhuri, H. R. & Majumdar S. (2006)

related to this phenomenon, especially among consumers in the USA and West Europe. $^{\rm 22}$

Social Structure	Primary Objects of Consumption	Drivers of Behavior		Principal Behavior Dimensions
Precapitalist - Feudal	Slaves, Women, Food	Military and Political Powers	Nobility	Pure ostentation
Modern - Capitalist	Very expensive products (e.g. Diamonds)	Social power, Status	Upper-middle	Ostentation, Signaling, Uniqueness
Post-Modern	Image, Experience	Self- expression, Self- image	Middle class, "Masses"	Uniqueness, Social conformation

Table 1: Conspicuous consumption during time

Source: Chaudhuri, H. R. & Majumdar S. (2006)

Chaudhuri & Majumdar propose a periodic analysis of conspicuous consumption behavior, "depicting its evolution, nature and character". Table 1 summarizes the main ideas in the paper, concerning changes in objects of conspicuous consumption, motives, dimensions and classes engaged in it, during time.

2.2 Reasons and Motivations of Conspicuous Consumption

In spite of the massive volume of research, the exact nature of the reasons behind conspicuous consumption and motivations that lead consumers toward it has not been fully investigated. Some of them are explained below, drawing upon relevant analyses made by researchers coming from different fields.

Sexual signaling (evolutionary viewpoint). De Fraja gives an interesting explanation for the universal human desire for increasing consumption, and the associated propensity of individuals to sacrifice survival activities (such as the acquisition of nutritious food, of adequate shelter, of health care, etc.), in order to buy other goods with zero or negative survival value (like luxury goods, leisure travel, entertainment, etc.).²³ He argues that this desire for conspicuous consumption was shaped during evolutionary times by a mechanism that biologists define as sexual selection. This occurs when an observable trait (in our case, conspicuous consumption) is used by members of one sex to signal their unobservable characteristics valuable to members of the opposite sex. Furthermore, it is explained that there is an analogy between the standard economics problem of utility maximization and the standard biology

²² Wisman, J. D. & Baker, B. (2011)

²³ De Fraja, G. (2008)

problem of the maximization of individual fitness (the ability to pass genes to next generations). A strict theoretical foundation is set up, to include conspicuous consumption in the utility function. Similarly, Pollet & Thienpont manage to show that conspicuous consumption has its roots in sexual selection, following Darwin's arguments and contrasting Veblen who stated that natural selection is unable to explain the persistence of "wasteful" behavior. They trace the origins of the phenomenon back to the theory of sexual selection, more particularly to the handicap principle.²⁴ According to these arguments, reliable and therefore costly signals (like conspicuous consumption) that signal a man's mate value will be developed, to avoid the risk of deception.

Drawing on several theories of different disciplines Sundie et al. made a thorough investigation of the nature of this behavior, by examining both the display (communication) and the perception (interpretation) sides of conspicuous consumption. The conducted experiments demonstrate that the motivation to conspicuously consume and display, in a mating context, is evident among men.²⁵ Despite this, observed increases in conspicuous consumption in response to mating motivations are not universal among them. For example, when explicitly primed with a desire for long-term mates, these unrestricted men did not increase in their conspicuous consumption tendencies. Furthermore, men who purchase luxury goods are perceived as more attractive specifically as short-term (but not long-term) partners. The research suggests that "conspicuous products such as Porsches might serve, for humans, an analogous function to that served by conspicuous tails for peacocks". Janssens et al. conducted other experiments with more than 100 male students participating, which further support the link between conspicuous consumption and male mating strategies.²⁶ Their study shows that exposure to a sexily dressed female experimenter increases single men's attention to status products (and decreases the attention to simple functional products) displayed to them.

Social signaling (emulation vs. conformity vs. uniqueness). Maybe the most effective demonstration of political and economic power of a ruler is the ability to consume part of the energy he controls for non-utilitarian purposes.²⁷ Hence, this conspicuous consumption is the most basic way in which the power of leaders can be symbolically reinforced. That is why monumental architecture and the consumption of luxury goods are a commonly understood expression of the power of leaders in previous civilizations, as explained above. Furthermore, this kind of behavior is known even to modern rulers, just assuming characteristics which are different from them of their predecessors.

According to the concept of "pecuniary emulation" introduced by Veblen in his theory of conspicuous consumption, individuals tend to imitate the consumption

²⁴ Pollet, T. & Thienpont, K. (2004)

²⁵ Sundie, J. M. et al. (2011)

²⁶ Janssens, K. et al. (2011)

²⁷ Trigger, B. G. (1990)

habits of other ones, situated at higher positions in the social hierarchy.²⁸ The members of each stratum try to emulate the consumption patterns of the class above it, in order to appear to be a member of the upper class. The result is that pressures to engage in conspicuous consumption affect even the poorest classes of the society.

Veblen evidenced also a second type of conspicuous consumption, related to the desire for uniqueness. Members of the upper class conspicuously consume, in order to differentiate themselves from individuals of lower classes.²⁹ Knowing that costs related to this kind of consumption must be large, this behavior can reduce imitation and further signal their high status. But, according to Veblen, this occurs less frequently than the pecuniary emulation. Individuals follow a specific consume pattern in order to be more like members of a higher income group, rather than seeking social distance from lower classes, because the former brings them a higher utility than the latter.

There is another motive that leads individuals toward conspicuous consumption, different from the two social needs mentioned above. This is the desire for conformity. Consumers try to conform to social norms, expectations and personal desires of other members of their own group.³⁰ Hence, individuals can engage in conspicuous consumption not only to imitate members of higher classes, but also to conform to other similar consumers, to stress their belonging to a certain group.

Auto signaling. There is a common characteristic in the motives mentioned in the previous section: the purpose of conspicuous consumption is oriented toward the society, toward other individuals in the same or different classes. A recent study finds out another category of reasons to engage in such a behavior, related to the consumer himself. Consuming luxurious visible products provides individuals with feelings of achievement, success and pleasure.³¹ Hence, conspicuous consumption is also a matter of self-image. This kind of consume can also be a means to fill social voids for members of marginal groups in the society.³² Therefore, conspicuous consumption may be in fact a compensatory consumption. Racial differences found in the behaviors related to conspicuous consumption, which are explained in the next section, confirm this idea.

2.3 Main Factors Influencing Levels of Conspicuous Consumption

A 2008's study conducted using USA representative data on consumption, show that Blacks and Hispanics spend larger shares of their total expenditure on conspicuous goods (clothing, jewelry, and cars) than do comparable Whites.³³ The observed differences are relatively constant over time, economically large and are associated

²⁸ Trigg, A. B. (2001)

²⁹ Amaldoss, W. & Jain, S. (2005b)

³⁰ Moawad, M.-H. (2007)

³¹ Moawad, M.-H. (2007)

³² Purinton, E. F. (2008)

³³ Charles, K. K. et al. (2008)

with substantial diversion of resources from other uses, such as health care and education. Another similar study used South African household data and found that non-White households spend relatively more (between 30 to 50 percent more) on visible consumption than do comparable White ones.³⁴ Even in this case, this implies lower spending on other consumption categories, especially on health and medical services. The reasons of these racial differences can be found in the discrimination and the marginal position of non-White individuals, even in several modern societies. Thus, this is a case of conspicuous consumption playing the role of compensatory consumption for non-White individuals.

Gender is another factor that may affect the level of conspicuous consumption, in two different and dualist ways. As mentioned above, conducted experiments demonstrate that the motivation to conspicuously consume and display, in a mating context, is evident among men.³⁵ While conspicuous consumption may serve as a mating strategy for men, mating conditions are irrelevant for women. Hence, we should find relatively higher levels of this consumption among men. On the other hand, due to the stronger social links between women, conspicuous consumption as a means of social signaling must be more present among them ³⁶.

Shukla suggests (based mainly on a survey conducted with customers of the BMW, Mercedes Benz and Lexus dealerships) that middle-aged consumers, like the youth segment, are a significant target group that needs to be studied in the area of conspicuous consumption. Middle-aged individuals have relatively higher incomes, a more stable career and a higher access to credit and debit resources, which give them higher capacities to pay for conspicuous consumption.³⁷ Levels of conspicuous consumption should also be higher among young and middle-aged consumers, due to the higher propensity to engage in social and sexual signaling. Even education may have a similar influence on the levels of conspicuous consumption. Higher levels of education generally bring higher access to financial resources and a stronger need to signal the achieved wealth, status, prestige, etc.³⁸ Examples coming from the cosmetics market confirm this idea, with the demand for conspicuous cosmetics that increases with price for college educated individuals and a normal downward-sloping demand curve for the ones who have not graduated. Thus, we should generally observe relatively higher levels of conspicuous consumption accompanying higher levels of education.

Finally, there is an interesting relationship between advertising and conspicuous consumption. A recent research paper formalizes the idea that advertising creates the possibility of conspicuous consumption, because it is a source of the signaling power of brands.³⁹ Advertisements inform the public of brand names and render them a

³⁴ Kaus, W. (2010)

³⁵ Sundie, J. M. et al. (2011)

³⁶ Amaldoss, W. & Jain, S. (2005a)

³⁷ Shukla, P. (2008)

³⁸ Amaldoss, W. & Jain, S. (2005b)

³⁹ Krähmer, D. (2005)

signaling device. In a price competition framework, the research shows that advertising increases consumers' willingness to pay for these products. But this can lead the firm selling to the "conspicuous consumer" to increase its expending on advertising. The later serve as an incentive to further increase levels of consume, and so forth…

3 Proposed Methodology - Conspicuous Consumption of the Albanian Households

As mentioned previously, there is an intended econometric approach upcoming. The intended study will be based on consumption data coming from the Living Standard Measurement Surveys Albania (LSMS), conducted by the Albanian Institute of Statistics (INSTAT) in 2002, 2005, 2008 and from the upcoming LSMS⁴⁰. Hence, the aim of this study is to make a cross-sectional analysis of the impact that "global" or "local" factors such as gender, age, education, etc. have on the level of conspicuous consumption of an Albanian family (working on data from the upcoming LSMS) and to find out changes during these time periods, if any. Therefore, a preliminary quantitative analysis was done, using the LSMS 2008 data. A similar sampling procedure to the one in LSMS 2005⁴¹ was carried out by INSTAT. It was considered a stratified two stage cluster sampling design in which the Primary Sampling Units (PSUs) were represented by the census Enumeration Areas (EAs), while the Second Stage Sampling units (SSUs) were the households (denoted as HUs). The EAs were stratified according to geographic criteria:

- Large geographic areas: "Mountain Area", "Coastal Area" and "Central Area" and their belonging to "Urban" area (big towns), "Other Urban" areas (i.e. small towns) and "Rural" areas.
- Tirana was considered as a separate stratum.

The LSMS final sample consisted of 3,600 households; 8 households selected for each of the 450 EAs selected at the first stage of the sampling. The selection of the EAs within each stratum was carried out by means of a Probability Proportional to Size (PPS) design; the measure of size was represented by the number of households living within each EA. The second stage units were selected by means of systematic sampling. In particular, within each selected PSU, 12 HUs were initially selected, 8 of them formed the base sample while the remaining 4 were considered as available substitutes. The selection of the new sample of 450 EAs has been carried out using the frame of EAs resulting from the Population Census at the end of editing stage. Before selecting the sample of PSUs, EAs from Tirana and Durrës have been quickly updated (quick counts) to take into account the migration flows.

⁴⁰ Even though it was planned to be conducted during 2011, there is a delay due to the implementation of the Albania Census 2011.

⁴¹ The World Bank (2006)

It is rather difficult to find in the relevant literature a clearly specified model showing the impact of different factors on the level of ostentatious consumption. On the other side, drawing upon this literature it is very easy to find theoretical explanations about these factors (some of them mentioned above) and their influence. We tried to include all these factors as independent variables and used a simple model of multiple linear regressions (estimated with the OLS method), with the monthly level of conspicuous consumption of the household as the dependent variable (including consumption in several "status-signaling" categories: domestic services; sports and hobby equipment, toys of all kinds and their repair; donations to church/mosque/non-profit institution; gambling and lottery losses, etc.) The obtained results are generally in accordance with expectations. But, it should be noted that the results contradict the theoretical viewpoint which argues that education is an incentive for the "desire for uniqueness" and consequently for ostentatious consumption, showing a negative effect of education (which is stronger for women). This may be attributed to a greater "consciousness" of the economic situation and a stronger "selfcontrol" among the highly educated individuals, behavioral aspects not previously studied in the relevant literature. An important conclusion is that the consumption of conspicuous goods is a small share of the total consumption of Albanian households. Gender and education of the household head are the factors which have the greater effects on the level of ostentatious consumption, at least among the influencing factors studied here (household size, stratum, gender and education of the head of the household, number of teenage members, etc.).

4 Conclusions

Conspicuous consumption is generally seen by researchers as a process or means to achieve or maintain status and other social benefits, mainly through the consumption of luxury goods. It is still almost impossible to give an exhaustive definition of it, maybe due to the interdisciplinary character of the phenomenon, which brings the need of multidisciplinary analysis combining viewpoints in economics, marketing, sociology, psychology and even biology. Conspicuous consumption is not a recent phenomenon. It can be found even in the antique civilizations and has always been present with periods of rising and falling down, related to social, economic and political background. Nowadays, this phenomenon is no more a privilege of the upper classes, but a common behavior even among the "masses". Conspicuous consumption is essentially an instrument of signaling. Individuals use it to signal sexual or social relevant characteristics to others. It can also be a matter of self-fulfillment or just serve as a means to fill social voids. It is because of this, that marginal groups spend relatively more on conspicuous consumption.

Higher levels of education should generally bring higher levels of such consumption, due to a higher access to financial resources and a stronger need to signal social characteristics. Similarly, levels of conspicuous consumption should also be higher among young and middle-aged consumers, due to the higher propensity to engage in social and sexual signaling. Women play an important role in the field of conspicuous consumption. It seems that their "presence" is important both for their part as direct consumers and as an incentive for the opposite gender to conspicuously consume. There is an interesting relationship between advertising and conspicuous consumption. The former increases consumers' willingness to conspicuously consume. This can lead the firm to increase its expending on advertising and therefore to cause a further increase in the level of conspicuous consumption, and so forth...

This research paper aims to be only the starting point of future more extended examinations in upcoming research. A questionnaire will be conducted (with students at the Faculty of Economy, University of Tirana) to find out which are the main reasons and motivations lying beneath conspicuous consumption of luxury goods in the Albanian environment. Moreover, an econometric analysis of the level of such consumption in the Albanian households will be developed, using data from the Living Standards Measurement Surveys (Albanian Institute of Statistics), for more than 3500 households. A preliminary quantitative analysis produced results that are generally in accordance with expectations. But, it should be noted that the results contradict the theoretical viewpoint which argues that education is an incentive for the "desire for uniqueness" and consequently for ostentatious consumption, showing a negative effect of education (which is stronger for women). This may be attributed to a greater "consciousness" of the economic situation and a stronger "self-control" among the highly educated individuals, behavioral aspects not previously studied in the relevant literature. An important conclusion is that the consumption of conspicuous goods is a small share of the total consumption of Albanian households. Gender and education of the household head are the factors which have the greater effects on the level of ostentatious consumption, at least among the influencing factors studied here (household size, stratum, gender and education of the head of the household, number of teenage members, etc.). However, it must be mentioned that the level of conspicuous consumption is always difficult to be measured, because individuals generally tend to not admit they are conspicuously consuming. Finally, conspicuous consumption has important implications, in microeconomic and macroeconomic levels, which has not been studied in this paper. These implications will be analyzed in upcoming research.

References

Amaldoss, W. & Jain, S. (2005a): Conspicuous Consumption and Sophisticated Thinking. *Management Science*. 51, 10, 1449-1466.

Amaldoss, W. & Jain, S. (2005b): Pricing Of Conspicuous Goods: A Competitive Analysis of Social Effects. *Journal of Marketing Research*. 42, 1, 30–42.

Bagwell, L. S. & Bernheim, B. D. (1996): Veblen Effects in a Theory of Conspicuous Consumption. *The American Economic Review*. 86, 3, 349-473.

Basmann, R. L., Molina, D. J. & Slottje, D. J. (1988): A Note on Measuring Veblen's Theory of Conspicuous Consumption. *The Review of Economics and Statistics*. 70, 3, 531-535.

Campbell, C. (1995): Conspicuous Confusion? A Critique of Veblen's Theory of Conspicuous Consumption. *Sociological Theory*. 13, 1, 37-47.

Charles, K. K., Hurst, E. & Roussanov N. (2008): *Conspicuous Consumption and Race*. URL: http://www.nber.org/papers/w13392.pdf, accessed on 05.08.2011.

Chaudhuri, H. R. & Majumdar S. (2006): Of Diamonds and Desires: Understanding Conspicuous Consumption from a Contemporary Marketing Perspective. *Academy of Marketing Science*. 2006, 11, 1-18.

De Fraja, G. (2008): *Sexual Selection and Conspicuous Consumption*. URL: www.le.ac.uk/economics/gdf4/originv1.pdf, accessed on 11.07.2011.

Evans, S. T. (2000): Aztec Royal Pleasure Parks: Conspicuous Consumption and Elite Status Rivalry. *Studies in the History of Gardens and Designed Landscapes.* 20, 3, 206–228.

Janssens, K., Pandelaere, M., Van den Bergh, B., Millet, K., Lens, I. & Roe, K. (2011): Can buy me love: Mate attraction goals lead to perceptual readiness for status products. *Journal of Experimental Social Psychology*. 47, 1, 254-258.

Kaus, W. (2010): *Conspicuous Consumption and Race: Evidence from South Africa*. URL: ftp://papers.econ.mpg.de/evo/discussionpapers/2010-03.pdf, accessed on 05.08.2011.

Krähmer, D. (2005): *Advertising and Conspicuous Consumption*. URL: http://www.sfbtr15.de/dipa/72.pdf, accessed on 05.08.2011.

Leach, H. (2003): Did East Polynesians Have a Concept of Luxury Foods?. *World Archaeology*. 34, 3, 442-457.

Marx, K. (1887): Capital: A Critique of Political Economy, Vol. I, Progress Publishers, Moscow, 404-427.

Moawad, M.-H. (2007): Les Facteurs Explicatifs de la Consommation Ostentatoire – Le Cas du Liban. URL: http://hal.archivesouvertes.fr/docs/00/41/39/21/PDF/these_mariehelene_moawad_sept_2007.pdf, accessed on 05.08.2011.

Murra, J. V. (1962): Cloth and its functions in the Inca state. American Anthropologist. 64, 4, 710-728.

Page, C. (1992): A History of Conspicuous Consumption, in Rudmin, F. W. & Richins, M. L. (eds.), *Meaning, Measure, and Morality of Materialism, Association* for Consumer Research, Kingston (Ontario), 82–87.

Patsiaouras, G. (2010): Rethinking Veblen's contribution to Consumer Research: a phenomenological enquiry into the perception of 'status consumption' by middle-income British consumers. URL: https://lra.le.ac.uk/handle/2381/8760, accessed on 22.06.2011.

Pollet, T. & Thienpont, K. (2004): Sexual Selection as A Mechanism for Conspicuous Consumption. URL: http://www.students.ncl.ac.uk/t.v.pollet/ISHE2004ThomasPolletKrisThienpont.pdf,

accessed on 11.07.2011.

Purinton, E. F. (2008): Compensatory or Conspicuous Consumption?: Bling It On!. URL: http://www.asbbs.org/files/2009/PDF/P/PurintonE.pdf, accessed on 05.08.2011.

Ray, N. M. (2009): *Household Consumption in Ancient Economies: Pompeii and The Wider Roman World*. URL: https://lra.le.ac.uk/handle/2381/8465, accessed on 22.06.2011.

Scheetz, T. (2004): A Modern Investigation of Status Consumption. URL: http://www.case.edu/artsci/dean/elf/presentationsreports.html, accessed on 02.07.2011.

Shukla, P. (2008): Conspicuous Consumption among Middle Age Consumers: Psychological and Brand Antecedents. *Journal of Product & Brand Management*. 17, 1, 25-36.

Sundie, J. M., Kenrick, D. T., Griskevicius, V., Tybur, J. M., Vohs, K. D. & Beal, D. J. (2011): Peacocks, Porsches, and Thorstein Veblen: Conspicuous Consumption as a Sexual Signaling System. *Journal of Personality and Social Psychology*. 100, 4, 664-680.

The World Bank (2006): Albania Living Standard Measurement Survey 2005 - BasicInformationDocument,http://siteresources.worldbank.org/INTLSMS/Resources/3358986-1181743055198/3877319-1190309366854/alb05bid.pdf, accessed on 02.02.2012.

Trigg, A. B. (2001): Veblen, Bourdieu and Conspicuous Consumption. *Journal of Economic Issues*. 35, 1, 99–115.

Trigger, B. G. (1990): Monumental Architecture: A Thermodynamic Explanation of Symbolic Behavior. *World Archaeology*. 22, 2, 119-132.

Veblen, T. B. (2007, [1899]): *The Theory of the Leisure Class*, Oxford University Press Inc., New York.

Wisman, J. D. & Baker, B. (2011): *Increasing Inequality, Inadequate Demand, Status Insecurity, Ideology, and The Financial Crisis of 2008.* URL: http://www.american.edu/cas/economics/pdf/upload/2011-1.pdf, accessed on 05.08.2011.

Market activity co-movements: Evidence from the S&P 500 index, crude oil and gold futures market

Michael Soucek1

¹ Chair of Finance and Capital Market Theory, European University Viadrina, Große Scharrnstr. 59, 15230 Frankfurt (Oder), Germany Email: soucek@europa-uni.de

Abstract. This study examines the co-movements between market activity on the equity, crude oil, and gold futures markets, proxied by open interest. The study is unique in its investigation of the simultaneous movements in the futures market liquidity. It provides empirical evidence that stock and crude oil futures market liquidity is positively related, but react negatively on sudden shocks in open interest. The level of instantaneous linkage reacts on external market condition. During periods of unstable financial markets, the correlation between equity and oil futures open interest decreases, and the correlation of the open interest on the equity and gold futures market becomes negative. This supports the hypothesis of gold being a natural hedging instrument of stock market fluctuation.

Keywords: Open Interest, Market co-movements, Market Liquidity

1 Introduction

Derivative markets play a crucial role in the price discovery process for publicly traded equities and commodities. In the last decade, futures trading has experienced enormous growth in all major asset classes. Current development raises many research questions, including co-movements between prices of different assets and commodities and their potential sources, as well as the nature of the relationship between their trading activity. The latter is the aim of this paper. Open interest, defined as the total number of contracts in place or open at the end of the trading day, proxies the demand for futures as hedging instruments and is often used as a proxy for market activity. Increasing open interest implies higher market activity. Aguenaou et al. (2011) argued that open interest is the real gain from the operations of the market concerned. The authors stated that the volume measures may obscure a clear understanding of the underlying demand through speculative intra-day trading being

included in the measure. Some researchers have used the open interest as a measure of market depth, heterogenous beliefs (Watanabe, 2001) and it is often used as a measure of futures and option liquidity. This study is unique in its investigation of simultaneous movements in open interest on the equity and commodity futures markets. Analyzing co-movements in the demand for hedging is crucial for understanding the source and nature of the equity and commodity market links and gives insights into the dynamics of market behavior. This study elaborates on the allocation of funds between various markets. Recently, numerous studies have examined the link between stock returns and commodity price changes. As one of the key input factors, oil in particular is assumed to determine equity prices worldwide.

Although prior academic research has delivered mixed results, the majority of studies agree on the significant co-movements between oil and equity prices. The price of gold, on the other hand, is assumed to be a good hedge for equity investors, as it has a very low correlation with equity prices. Its role for diversifying an investor's portfolio has been investigated widely and inclusion of gold holdings has been advocated to lead to a more balanced portfolio by improving its risk return characteristics (Ratner and Klein, 2008). Commonly, the relationship between oil prices and gold prices is assumed to be positive, implying that oil and gold are to some extent substitutes as hedging instruments (Zhang and Wei, 2010; Sari et al., 2010). Very recent studies by Büyüksahin and Robe (2010, 2011) have shown that the level of the relationship between commodity and equity markets has grown over the last decade due to a rising number of cross-market traders and hedge funds participating in the market. The authors have documented that in addition to macro-economic fundamentals, variations in the composition of the open interest in the commodity futures markets indeed help explain fluctuations in the extent of commodity-equity co-movements.

The aim of this paper is to investigate market activity co-movements; that is the link between the demand for hedging on various markets proxied by the open interest. It examines the lead-lag relationship between the open interest of the S&P 500 index, light sweet crude oil, and gold futures traded on the New York Mercantile Exchange (NYMEX) and the pattern of their dynamic correlation over the last 14 years. The key questions this paper tackled are first, do market activity changes on one market cause changes in the activity on other markets? Second, how does the demand for hedging react to a sudden shock in market activity on other futures markets? Third, this paper investigates contemporaneous time-varying correlation between market activity, using the DCC-GARCH framework by Engle (2002) to capture the level of market integration and its response to exogenous information shocks. The results help us to understand the source of cross-market linkages and liquidity flow on the financial markets.

The rest of the paper is organized as follows: Section 2 reviews the literature on crossmarket linkages and open interest. Sections 3 and 4 present the data and methodology used. Section 5 presents the empirical results, and section 6 concludes the paper.

2 Literature review

The impact of changes in oil price on equity prices is still a relatively young field of research, but particularly in the last few years, the number of papers on this topic has increased remarkably. The pioneering paper by Jones and Kaul (1996) tested the impact of oil prices on four developed markets using the cash-flow dividend model, finding a significant negative correlation with the U.S. stock market. The results for the Japanese and U.K. markets were inconclusive. On the other hand, most of the recent studies have supported the hypothesis that a significant link is present between numerous national stock indices and oil price time series (e.g. Driesprong et al., 2008; Basher and Sadorsky, 2006; Hammoudeh and Li, 2005). Many scholars have examined the nature and influences of the gold market (Tully and Lucey, 2007) and their linkage to various macroeconomic variables (Sari et al., 2010). Concerning stock market movements, gold returns are assumed to have a very low correlation with the stock market, and gold is meant to be used as a stock volatility hedge in portfolios (Sherman, 1983; Chua et al., 1990). Extended literature has also shown the link between gold and the prices of other commodities. Sari et al. (2010) found a weak and asymmetric relationship between oil and gold prices. Zhang and Wei (2010) reported a long-term cointegrating price link, with their results indicating oil price linearly Granger causes the gold returns. Overall, the relationship between oil price changes and equity returns is assumed to be positive.

The number of studies on the economic relevance of open interest is still assessable. This variable is usually used to account for liquidity or demand for hedging and also to judge market and speculation activity. Chang et al. (2000), for instance, discussed the case of the S&P 500 stock index futures and showed that in the periods of high volatility, open interest for large hedgers tends to increase. The authors implied that speculation is surprisingly not that strong in periods of high volatility. Considerable amount of literature has examined the link between futures prices and the underlying open interest. Yang et al. (2004) investigated the long-run relationship and the role of open interest for futures prices. Their findings showed futures prices causing open interest, but not vice versa in the long run. On the contrary, the results presented by Wang and Yu (2004) showed that futures portfolio returns are negatively related to lagged changes in open interest and contradicted the hypothesis of efficient markets. The idea of analyzing the linkage between the open interest on different futures markets is rather new. Aguenaou et al. (2011) analyzed simultaneous determinants of the open interest in cross-listed settings. They were the first to examine how the open interest pertains to the interaction between two futures markets, the EUREX and the LIFFE. Their study implied that an external shock strongly influences the interaction between the two markets and provided evidence for causalities between the open interests on the futures markets.

3 Data and Methodology

3.1 Data and preliminary analysis

This paper compares market activity for three of the major futures contracts in the U.S.: the stock index S&P 500, light sweet crude oil, and gold. Futures data are considered for the period from January 1998 to July 2011 and are sourced from Datastream. The S&P 500 futures are traded in New York on the open outcry and on the CME Globex. The delivery months are quarterly in March, June, September, and December, with the last trade day usually being the third Thursday of every contract month. All positions are settled in cash. The contract size for the S&P 500 index future is \$ 250 the futures price. The considered oil contract is the light sweet crude oil futures traded on the NYMAX. They are listed nine years forward and traded almost around-the-clock. This is also the case for the considered gold futures quoted on COMEX. Their trading is conducted for delivery during the current calendar month; the next two calendar months; any February, April, August, and October falling within a 23-month period; and any June and December falling within a 72month period beginning with the current month. Arouri and Nquyen (2010) argued that for returns, weekly data may capture the interaction price changes better than daily or monthly data. The use of weekly data in the analysis instead of daily data significantly reduces any potential biases that may arise such as the bid-ask effect, non-synchronous trading days, etc. Typically, more information arrives on financial markets in the middle of the week, which should be captured if open interest is assumed to be a function of demand for hedging (Aguenaou et al., 2011; Chang et al., 1997). As a result, in this paper the open interest log changes have been calculated on Wednesday. We consider a full period sample and analyze the number of all futures contracts on a particular week day.

	S&P	Oil	Gold	Δ S&P	Δ Oil	Δ Gold
Obs	705	705	705	705	705	705
Mean	13.14	13.56	12.48	-0.05	0.19	0.15
Median	13.23	13.47	12.53	0.91	0.69	0.13
SD	25.88	45.31	48.29	8.08	4.11	5.21
Kurt.	2.00	1.42	1.86	14.45	4.24	4.07
Skew	-0.44	0.14	-0.03	-2.93	-0.73	0.05
AR(1)	0.95	0.99	0.99	-0.26	-0.12	0.17
Q(1)	0.00	0.00	0.00	0.00	0.00	0.00
JB	0.00	0.00	0.00	0.00	0.00	0.00

Table 1: Descriptive statistics

Note: Descriptive statistics is given for weekly logarithmic open interest and logarithmic changes in the open interest. Q(1) is the 2 distributed Box-Ljung test statistics on autocorrelation. JB stands for the Jacque-Bera test-statistics with the null hypothesis that the variable is normally distributed

The data presented in Table 1 consider 705 weekly observations for the U.S. market. The changes in open interest do not follow normal distribution, since the null hypothesis of the Jacque-Bera test can be rejected on the 1% level for all time series. The data for the S&P 500 futures exhibit the highest volatility and asymmetry within the data examined. One of the reasons for this finding might be the smaller number of futures with different maturities traded at the same time compared to the commodity contracts. All series in levels and in first differences exhibit significant first-order autocorrelations. The variables are linearized. An average of 13.23 is equal to around 556,800 open futures contracts 10 at the end of the trading day.

The stationarity of the observed univariate time series is the key assumption of the OLS-based VAR(p) estimation, tests for Granger causality, and DCC-GARCH estimation. In the following, the Augmented Dickey-Fuller (ADF) test for unit roots and KPSS stationarity test are adopted. The results for both well known approaches are presented in Table 2.

	ADF		KPSS Test	
Future	I(0)	I(1)	I(0)	I(1)
S&P 500 ln(OI)	-2.80	-25.62*	2.25*	0.20
Oil ln(OI)	-2.73	-25.90*	9.26*	0.05
Gold OI	4.02*	-19.32*	8.72*	0.05

Table 2: Unit root test

Note: Open interest is given in logs. ADF is the Augmented Dickey-Fuller test. KPSS stands for the Kwiatkowski-Phillips-Schmidt{Shin test. The t-values are compared with the Dickey-Fuller (DF) critical values. *denotes signicance at 1% level. Remaining test-statistics are not signicant at 10% level.

In some cases the log open interests seem to be stationary over the observation period, nevertheless the first differences in the log open interest are stationary for all time series. This result allows to employ analysis based on standard vector autoregressive framework as well as the multivariate GARCH specification for linearized open interest changes.

3.2 Methodology

The first part analysis of the open interest spillover is determining the lead-lag relation between the time series and is based on simple vector autoregressive model used for instance by Sims (1980). In the basic form, a VAR consists of a set *K* endogenous variables. Let be a vector with $\mathbf{y}_t = (y_{1t}, \dots, y_{kt}, \dots, y_{Kt})$ for $k = 1, \dots, K$. The VAR(*p*)-process is then defined as:

$$\mathbf{y}_{\mathbf{t}} = A_1 \mathbf{y}_{\mathbf{t}} + \ldots + A_p \mathbf{y}_{\mathbf{t}-\mathbf{p}} + \mathbf{u}_{\mathbf{t}},$$

where A_i is a $K \times K$ coefficient matrix and \mathbf{u}_t is a K - dimensional white noise process. To obtain stable results, the underlying time series has to be stationary. In following, underlying VAR(P) model is exploited to test Granger causality relationships and orthogonalized impulse response functions.

Second part of the paper evaluates the time-varying conditional correlation between the changes in the open interest using the DCC-GARCH model of Engle (2002). We consider basic multivariate GARCH model for changes in open interest \mathbf{y}_t

$$\mathbf{y}_t = ER(\mathbf{y}_t \mid \mathbf{F}_{t-1}) + \boldsymbol{\varepsilon}_t$$

with

$$\varepsilon_t = D_t \eta_t$$
 and $Q_t = D_t \Gamma_t D_t$

The changes in the open interest conditional to the information set F_{t-1} of the time series are assumed to be normally distributed

$$\mathbf{y}_t | \mathbf{F}_{t-1} : N(0, Q_t) \quad t = 1, 2...T,$$

where $D_t = diag(h_1^{1/2}, ..., h_m^{1/2})$ is a diagonal matrix of conditional variances. The conditional variance h_{it} is defined as a univariate GARCH(p,q) process:

$$h_i t = \omega_i + \sum_{r=1}^p \alpha_{ir} \varepsilon_{i,t-r}^2 + \sum_{s=1}^q \beta_{is} h_{i,t-s}$$

If the η_t is a i.i.d. vector of random variables, with zero mean and unit variance, the Γ_t is the conditional correlation matrix of the standardized residuals $\eta_{it} = y_{it}/\sqrt{(h_{it})}$. The dynamic conditional correlation can be estimated as follows:

$$T_t = \{ diag(Q_t) - 1/2 \} Q_t \{ diag(Q_t) - 1/2 \},\$$

where $k \times k$ is symmetric positive definite matrix Q_t given by

$$Q_t = (1 - d_1 - d_2)Q + d_1\eta_{t-1}\eta_{t-1}' + d_2Q_{t-2}.$$

 \overline{Q} stands for unconditional variance matrix of the η_t . The parameters d_1 and d_2 are scalars which capture the effect of previous shocks and previous dynamic correlation on the current conditional correlation. They have non-negative value and should satisfy $d_1 + d_2 < 1$, which implies Q_t being greater than zero. Equation 7 is a conditional covariance matrix and \overline{Q} is the unconditional variance matrix of η_t . DCC is not linear but can be estimated using two-step method based on classical maximum likelihood function. First step is to estimate the univariate GARCH from equation 5 and second step is the estimation of correlation coefficients. For technical details about the estimation the reader is referred to Teräsvirta (2009).

4 Empirical Results

4.1 Granger causality test

The first part of the analysis is based on the VAR(p) model presented in section 4. Analyzing the estimated coefficients themselves sometimes reveals uninteresting results. VAR models are used, in most cases, to test Granger causality and to determine whether lags of one variable help explain the current value of some other variables (Granger, 1969). Because the first differences in the logarithmic open interest are stationary, it is permissable to use the framework described above. Let us assume a VAR(p) model from section 4. y_k does not Granger cause y_1 if all of the coefficients on lagged values of y_k are zero in the equation for y_1 . This linear restriction can be tested using Wald-statistics (Lütkepohl, 2005). The lag length of the underlying VAR model is based on SIC. Because the relationship is not symmetric, Granger causality in one direction does not imply causality in the other direction. The data in Table 3 summarizes the multi-equation F-test for testing the equality of coefficients. The first column represents the endogenous variables in a single equation from the VAR(p) system. The values are test statistics for the variables in the row to Granger cause the variable in the column. For instance, the F-value that the S&P 500 index futures open interest Granger causes the open interest of the light sweet crude oil future is 8.36, which is signifcant at the 1% level. It seems, that a bidirectional causality exists between the liquidity on the S&P 500 futures market and the oil futures market. Furthermore, the changes in the open interest of the gold future on the COMEX has a significant impact on the future movement in the open interest of light sweet crude oil futures. This result might be surprising, since Zhang and Wei (2010), for example, showed that oil prices tend to Granger cause the gold price. It would be reasonable to assume that market activity would follow the same pattern, because a significant relationship exists between futures prices and open interest (Yang et al., 2004; Wang and Yu, 2004). Columns 5 and 6 in Table 3 test for the block exogeneity; that is, the significance of the variable in the first row for the system of the two remaining equations together. Column 6 is testing the instantaneous causality of a variable for the system. If in period t, adding $y_{k,t+1}$ to the information set improves the forecast of $y_{1,t+1}$, an instantaneous causality exists between the two variables (Lükepohl, 2005).

	Multivariate Granger caus.			Block Exogenity	
	S&P 500	Oil	Gold	GC	Inst.
					causality
S&P 500	-	8.36	0.012	5.27	20.7
		(0.00)	(0.91)	(0.00)	(0.00)
Oil	4.59		1.23	4.35	21.47
	(0.03)	-	(0.27)	(0.00)	(0.00)
Gold	0.39	7.8	-	1.92	3.64
	(0.53)	(0.01)		(0.05)	(0.16)

Table 3: Multivariate Causality F-Test

Note: First three columns give the Granger causality t-values for pairs of equations within the system. The first column represents endogenous variables in a single equation from the VAR(p) system. The values are F-statistics for the variables in the row to Granger-cause the variable in the column. The columns 5 is test for the block exogenity; i.e. the significance of the variable in the first row for the equation system. Column 6 is testing the instantaneous causality of a variable for the system. The lag length of the underlying VAR model is based on SIC.

4.2 Impulse response functions

Impulse response functions are dynamic simulations showing the response of an endogenous variable over time to a given shock. Figures 1 to 3 illustrate the impulse responses resulting from an one standard deviation shock in the open interest changes for the analyzed S&P 500, light sweet crude oil, and gold futures traded in New York. Monte Carlo constructed 95% confidence bounds are provided to judge the statistical significance of the impulse response functions. Figure 1 shows that a shock in the market activity on the stock futures market has a negative and statistically significant initial impact on the open interest of oil futures on the NYMEX. The impact remains significant for more than three weeks into the future. The shock on the stock index futures markets, nevertheless, does not have an impact on the number of the gold futures contracts opened on the COMEX. The response on an one standard deviation shock in the open interest of light sweet crude oil futures is significant for the future trading activity of both the gold and the S&P 500 index futures. A shock in the gold futures open interest on the COMEX does not have an impact on the trading activity on the other two futures markets. The calculations support previous results on the significant relationship between the equity and oil futures market liquidity. The gold futures open interest seems to be unrelated to the equity market, but exhibits a certain level of linkage to the oil market liquidity. Shocks in the oil market activity seem to have an impact on gold futures trading.

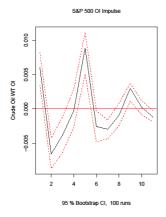
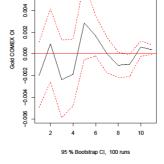


Figure 1: Impulse response function S&P 500 open interest shock

0.006

(a) Crude oil futures OI response



S&P 500 OI Impulse

(b) Gold futures OI response

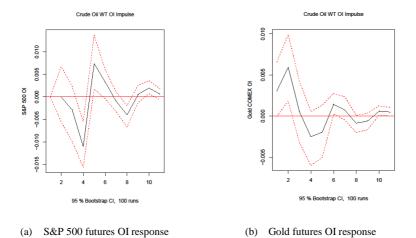
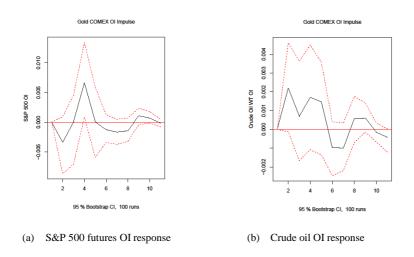


Figure 2: Impulse response function of Light Crude Oil NYMEX Future open interest shock

Figure 3: Impulse response function of Gold Future open interest shock



4.3 Conditional correlations

This last empirical section investigates the pattern of the time-varying correlation between market activity on the gold, crude oil, and S&P 500 futures market. The key

question, furthermore, is whether a co-movement exists in the market activity between the financial and commodity markets. The calculations is based on DCC-GARCH(1,1) model by Engle (2002). The estimates of the time varying conditional correlations are plotted in figure 3. The estimated parameters and introduced in section 4 are $d_1 = 0.024$ and $d_2 = 0.921$; i.e. they have non-negative value and satisfy the condition $d_1 + d_2 < 1$.

The conditional correlation between the S&P 500 index future and the light sweet crude oil future open interest is plotted in the Figure 4(a). Generally, we observe a moderate positive correlation between the trading activity on the two markets. The level of the correlation is similar to the level generally estimated for price returns (Chang et al., 2010); i.e., between 0.2 and 0.4. The results imply that rising market activity measured with open interest on the stock index futures market is related to an increase in the activity on the market for crude oil futures. This observation speaks for an advanced level of integration between the markets. The investors do not flee stocks and move to oil, which would result in a negative correlation. Stocks and oil are known to be related to general economic variables, with favorable economic conditions resulting in higher prices. The nature of the trading activity co-movement on the two markets seems to be related to the general economic environment more than to resource allocation between the trading places and hedging. Nevertheless, two major structural breaks exist in the time-varying correlation curve. The first begins around 2001 and ends in 2003. During this period, the correlation falls far under 0.1, indicating the absence of a trading activity linkage between the two markets. Between 2001 and 2003, financial markets suffered from the consequences of the bursting dot.com bubble, which led the markets to exhibit high volatility. This was followed by widespread economic recession. Furthermore, between 2001 and 2003, additional external events occurred, which had a significant impact on the volatility and link between oil and stock prices. These events included the 9/11 terrorist attacks in 2001 and the start of the Iraq war in March 2003 (Filis et al., 2011). The trading activity comovement between equity and crude oil futures at these times was very weak. The demand for hedging on the markets appeared to be driven by different forces. In other words, there is no clear structure in the market activity allocation between oil and equity futures markets. We observe a similar reaction during the global economic downturn in 2008 and early 2009. Between 2004 and 2007, the correlation varies around 0.3, even though the market activity link experienced an enormous decline in 2008. If we look more closely at the Figure 4(a), surprisingly, the decline is not a reaction to the bankruptcy of Lehman Brothers, which is usually considered to be the start of the financial crash, but have already occurred in the middle of 2008. Between March and July 2008, the first contours of the upcoming financial crises were noticeable. The Dow Jones Industrial Average reached the two year minimum and Bear Strearns bank was acquired by JPMorgan Chase in a deal backed by the Federal Reserve. Nevertheless, oil prices continued to rise until July 2008. The markets were not expecting the economic recession yet. The global market downturn starting in September 2008 allowed the correlation between market activity to rise again. This supports the assumption that the connection between the two markets is related to certain external economic conditions, since in autumn 2008 the financial and banking crises grew into the crises affecting the entire global economic system. Figure 4(b)

illustrates the time-varying correlation of the open interest of the S&P 500 futures and gold futures. Generally, trading activity on the gold market seems to be unaffected by the open interest of the equity futures, fluctuating around zero or having a weak negative correlation. Nevertheless, we again observe structural changes in periods of financial distress. In both periods of stock market turbulence, the DCC curve exhibits a significant decline, implying a negative dynamic correlation. This supports the hypothesis that gold is a hedging instrument for equity investors mentioned, for instance, by Sherman (1983). Whereas the trading activity of S&P 500 futures is decreasing, part of the volume is allocated to the gold market, and the gold futures are related to open interest growth. The investors flee stock futures and move to gold futures market.

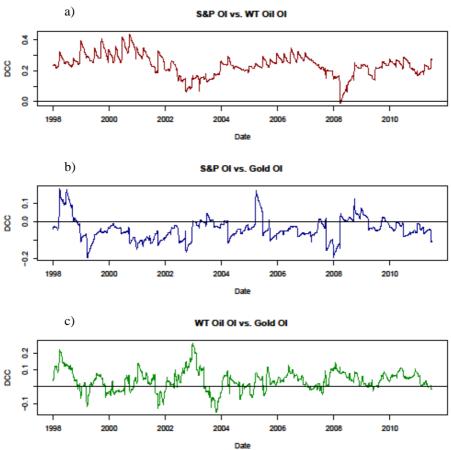


Figure 4: Dynamic conditional correlation (DCC)

Note: Time-varying correlation between the open interst of the S&P 500, Light Crude Oil and Gold futures. The estimates are based on DCC-GARCH(1,1) model by Engle (2002).

Figure 4(c) presents the dynamic conditional correlation between the open interest of the light sweet crude oil and gold futures on the NYMEX. This figure supports the observation of the relationship between the stock and commodity market activity. Because the correlation is mostly fluctuating around zero, a strong positive peak is observed between 2002 and 2003. This is inline with the decreasing correlation between equity and crude oil open interest and even the negative correlation between equity and gold market activity. A possible explanation might be that in periods of less volatile markets and economic upturn, trading activity with equity and oil futures is driven by the favorable economic environment and is not used as a hedge for stock price movements. This cross-market relationship changes in periods of economic downturn, when oil stops following the activity of the equity markets and exhibits a stronger linkage to the gold market.

Overall, the empirical evidence suggests a link between market activity on the equity and commodity futures markets in the U.S. Rising open interest on the equity market is related to rising trading activity with crude oil futures. The level of integration of these two markets drops dramatically during periods of financial and economic distress, such as the economic recessions of 2002 and 2008. During these periods, the correlation between open interest for the equity futures and gold futures tends to be negative. This speaks to the value of investors allocating their wealth in safe assets such as gold, whereas in more quiet period the correlation fluctuates around zero.

5 Conclusion

This study investigates simultaneous movements in trading activity on the equity and commodity futures markets. It evaluates the causality relationships between open interest of the S&P 500, light sweet crude oil, and gold futures, as well as impulse response functions based on shocks in the open interest change. The findings indicate a significant bi-directional causality between market activity on the equity and crude oil futures markets. Additionally, gold futures open interest Granger cause oil demand for hedging. Impulse response functions support these results. A shock in the equity futures open interest is immediately followed by a negative response of the crude oil futures open interest and vice versa. A negative shock lasts for about four weeks. The gold futures open interest changes react positive based on the oil market activity shock. The main content of this paper is the analysis of the dynamic correlation pattern based on DCC-GARCH(1,1) by Engle (2002).

In summary, this study provides evidence for integrating market activity between the commodity and equity futures markets. In particular, the link between light sweet crude oil and S&P 500 futures seems to be strong. While the dynamic correlation is mostly positive, it appears that a positive shock on one of the two markets is followed by a negative reaction on the other. The stock and oil markets are assumed to have a considerable level of integration in the price discovery process. This study provides empirical evidence that significant interdependencies exist in market activity as well. Looking at the shape of DCC-correlation curve, two major results are worth mentioning. First, the correlation between equity and oil trading activity is positive.

Second, the level of cross market liquidity integration reacts to external information shocks in the economic environment. During the financial crisis periods in 2002 and 2008, the linkage between the oil and equity futures market liquidity disappeared and, equity and gold futures exhibit a negative correlation. This is interpreted as empirical evidence for investment allocation during periods of financial and economic distress. Market participants move their money toward more save investments. Oil on the other hand, decreases its link to the equity market activity.

For further research, it would be interesting to extend the model to consider macroeconomic variables. One might also consider other proxies for liquidity or market activity and extend this study to various markets. It will be also necessary to consider different time intervals for measuring liquidity. Several studies have exploited the relationship between financial and commodity markets. This study extends the literature and intends to call the attention to the nature and source of the market integration as a result of global asset allocation.

References

Aguenaou, S., Ap Gwilym, O., Rhodes, M. (2011) Open interest, cross listing, and information shocks. *Journal of Futures Markets*. 31, 755-778

Aloui, C., Jammazi, R. (2009) The effects of crude oil shocks on stock market shifts behaviour: A regime swiching approach. *Energy Economics.* 31, 789-799

Arouri, M. E. H., Nquyen, D. K. (2010) Oil prices, stock markets and portfolio investent: Evidence from sector analysis in europe over the last decade. *Energy Policy*. 38, 4528-4539

Basher, S. A., Sadorsky, P. (2006) Oil price risk and emerging stock markets. *Global Finance Journal*. 17, 224-251.

Büyüksahin, B., Robe, M. A. (2010) Speculators, commodities and cross-market linkages. *Working Paper*. (November 28, 2010). Available at SSRN: http://ssrn.com/abstract=1707103

Büyüksahin, B., Robe, M. A. (2011) Does paper oil matter? energy markets financialization and equity-commodity co-movements. *Working Paper*. (July 28, 2011), Available at SSRN: http://ssrn.com/abstract=1855264

Chang, C.-L., McAleer, M., Tansuchat, R. (2010) Conditional correlations and volatility spillovers between crude oil and stock index returns. *Discussion Paper*. *Kyoto Institute of Economic Research, KIER Discussion Paper Series*, 715-758

Chang, E., Chou, R. Y., Nelling, E. F. (2000) Market volatility and the demand for hedging in stock index futures. *Journal of Futures Markets*. 20, 105-125

Chang, E. C., Pinegar, M. J., Schachter, B. (1997). Interday variations in volume, variance and participation of large speculators. *Journal of Banking and Finance*. 21, 797-810

Choi, K., Hammoudeh, S. (2010) Volatility behavior of oil, industrial commodity and stock markets in a regimeswitching environment. *Energy Policy*. 38, 4388-4399

Chua, J., Sick, G., Woodward, R. S. (1990) Diversifying with gold stocks. *Financial Analyst Journal*. 46, 76-79

Driesprong, G., Jacobsen, B., Maat, B. (2008) Striking oil: Another puzzle? *Journal of Financial Economics.* 89, 307-327

Engle, R. (2002) Dynamic conditional correlation. A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*. 20, 339-350

Ferris, S. P., Park, H. Y., Kwangwoo, P. (2002) Volatility, open interest, volume, and arbitrage: Evidence from the S&S 500 futures market. *Applied Economics Letters*. 9, 369-372

Filis, G., Degiannakis, S., Floros, C. (2011) Dynamic correlation between stock market and oil prices: The case of oil-importing and oil-exporting countries. *International Review of Financial Analysis.* 20, 152-164

Gisser, M., Goodwin, T. (1986) Crude oil and the macroeconomy: tests of some popular notions. *Journal of Money, Credit and Banking*. 18, 95-103

Granger, C. W. J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*. 37, 424-438

Hammoudeh, S., Li, H. (2005) Oil sensitivity and systematic risk in oil-sensitive stock indices. *Journal of Economics and Business*. 57, 1-21

Hong, H. (2000) A model of returns and trading in futures markets. *Journal of Finance*. 55, 959-988

Hooker, M. (1996) What happened to the oil price-macroeconomy relationship? *Journal of Monetary Economics.* 38, 195-213

Jones, C., Kaul, G. (1996). Oil and sthe stock markets. Journal of Finance. 51, 463-491

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal Econometrics*. 54, 159-178

Lee, K., Ni, S., Ratti, R. (1995) Oil shocks and the macroeconomy: The role of price variability. *Energy Journal*. 16, 39-56

Lütkepohl, H. (2005). New Introduction to Multiple Time Series Analysis, Springer, Berlin.

Malik, F., Hammoudeh, S. (2007) Shock and volatility transmission in the oil, us and gulf equity markets. *International Review of Economics & Finance*. 16, 357-368

Mork, K. (1989) Oil and the macroeconomy when prices go up and down: An extension of Hamilton's results. *Journal of Political Economy*. 97, 740-744

Ratner, M., Klein, S. (2008) The portfolio implications of gold investments. *The Journal of Investing*. 17, 77-87

Sari, R., Hammoudeh, S., Soytas, U. (2010) Dynamics of oil price, precious metal prices and exchange rate. *Energy Economics*. 32, 351-362

Sherman, E. J. (1983) A gold pricing model. *Journal of Portfolio Management.* 8, 68-70

Silvennoinen, A., Teräsvirta, T. (2009) Multivariate GARCH models, in: T. G. Anderson, R. A. Davis, J.-P. K., Mikosch, T. (eds), *Handbook of Finacial Time Series*, Springer, New York, 201-229

Sims, C. A. (1980) Macroeconomics and reality. Econometrica. 48, 1-48

Tully, E., Lucey, B. M. (2007) A power garch examination of the gold market. *Research in International Business and Finance*. 21, 316-325

Wang, C., Yu, M. (2004) Trading activity and price reversals in futures markets. *Journal of Banking and Finance*. 28, 1337-1361

Watanabe, T. (2001) Price volatility, trading volume, and market depth: Evidence from the Japanese stock index futures market. *Applied Financial Economics*. 11, 651-658

Yang, J., Bessler, D. A., Fung, H.-G. (2004) The informational role of open interest in futures markets. *Applied Economics Letters*. 11, 569-573

Zhang, Y.-J., Wei, Y.-M. (2010) The crude oil market and the gold market: Evidence for cointegration, causality and price discovery. *Resources Policy*. 35, 167-177

Factors influencing bank profitability – comparison of banks in Serbia, Croatia, Slovenia and the 10 largest banks in EU area

Ana Polovina¹, Ljiljana Popovic²

¹ Ph.D. Student, Department of Industrial Engineering and Management, University of Novi Sad *polovina.ana@gmail.com*

² Ph.D. Student, Department of Industrial Engineering and Management, University of Novi Sad *ljiljana.popovic.ns@gmail.com*

Abstract. The purpose of this research is to examine the relationship between bank-specific, market-specific and macroeconomic factors and bank profitability in three countries that all went through transition (Serbia, Croatia and Slovenia) and the 10 largest banks in EU area. In that respect, individual bank data, market-specific and macroeconomic data were collected for the 2004-2010 period. Fixed effects regression model will be used for estimating each bank's profitability, which is measured by return on average assets (ROAA), and it will be tested with the Hausman test. Considering that these three countries have gone through the process of transition at different periods and are now in different positions when it comes to EU membership, research results should reveal the differences in factors that influence profitability of banks in these countries compared to the 10 largest EU banks.

Keywords: banking, profitability, ROAA, transition countries

1 Introduction

Developments in banking sector affect not just a banks, but economy of the country as a whole. Stable and successful banking sector is of interest to management and stakeholders, as well as to employees, customers and government authorities. The better a bank's profit situation is, the better its prospects are for issuing new capital, because investors have greater confidence in the bank's financial strength (Rumler and Waschiczek, 2011).

Therefore, the analysis of factors influencing bank's profitability is essential to management of banks and it helps in making correct business decisions.

The purpose of this paper is to investigate the influence of bank-specific, market-specific and macro-economic factors on bank profitability in the former Yugoslavia countries – Serbia, Croatia and Slovenia and the 10 largest banks in EU area, for the period 2004-2010.

Since it was difficult to collect data for all commercial banks in EU area, authors have analyzed the 10 banks that have largest total assets in the European Union, at 31 December 2010 (BNP Paribas, HSBC Holdings, Barclays, Credit Agricole S.A., Royal Bank of Scotland, ING Group, Santander Group, Llioyds Banking Group and Société Générale).

2 Literature Review

A large number of studies have already dealt with the determinations of bank profitability. Some of them have focused on a particular country and some of them on a panel of countries. Studies on the determinations of bank's performance in the US were carried out by Berger (1995), Angbazo (1997) and Neeley and Wheelock (1997). Some other studies have focused on emerging countries – Brasil (Afanasieff et al., 2002), Tunisia (Ben Naceur and Goaied, 2001), Taiwan (Ramlall, 2009) and Pakistan (Gul, Irshad and Zaman, 2011). Most of the panel countries were focused on European banks (Molyneux and Thornton, 1992, Abreu and Mendes, 2002 and Pasiouras and Kosmidou, 2007). Molyneux and Thornton (1992) find a significant positive relation between the return on equity and the level of interest rates, bank concentration and government ownership in the research based on 18 European countries during the period of 1986-1989. Abreu and Mendes (2002) report that well capitalized banks face lower bankruptcy costs and this advantage "translate" into better profitability. Their research, also, show that the unemployment rate and inflation rate are relevant in explaining bank profitability. The study of Pasiouras and Kosmidou (2007) shows that profitability of both domestic and foreign banks in EU area is affected not only by bank's specific characteristics but also by financial market structure and macroeconomic conditions.

There are also some studies which examine the influence of different factors on bank profitability in the South-Eastern European area countries (Kosak and Cok, 2008, Pejic Bach, Posedel and Stojanovic, 2009 and Athanasoglou, Delis and Staikouras, 2006). Kosak and Cok (2008) show the positive relation between capital strength, cost efficiency and credit risk exposure and performance indicators. Also, negative relation between liquidity management and bank asset structure and performance indicators is shown. Pejic Bach, Posedel and Stojanovic (2009) focus their research on the profitability of banks in Croatia in the period 1999-2005. The results show that in short term and in stable macro-economic conditions, higher profitability is accomplished by well-capitalized banks with larger market share, which are also efficient in managing their operating costs and able to pass "non-earning assets" costs to their clients. Athanasoglou, Delis and Staikouras (2006) find that market share is positively correlated with bank profitability and inflation has a strong effect on profitability, while bank profits are not significantly affected by real GDP per capita fluctuations (probably owing to the small sample period).

However, according to the knowledge of the authors, the research based on the comparison of factors influencing profitability of banks in the transition countries in

South-Eastern European area and banks in EU area has not been conducted yet. Therefore, the aim of this research is to identify differences in factors influencing bank's profitability in countries that have gone through a transition (Serbia, Croatia and Slovenia) and the 10 largest banks in EU area.

3 Data and Methodology

This study includes data on the profitability of the banking business and the macroeconomic indicators in the period of 2004 to 2010. Bank-specific data were collected from the individual financial statements of banks (balance sheets and income statements). Country and market specific data, such as inflation and real GDP growth, were obtained from statistical publications of central banks and European statistics database (Eurostat). Data on market capitalization were collected from Standard & Poor's and Global Stock Markets Factbook database. The panel dataset consists of 31 banks in Serbia, 32 banks in Croatia, 15 banks in Slovenia and 10 largest banks in EU area. All the data are on an annual basis. Factors that have influence on profitability of banks are split into three categories: bank-specific variables, market-specific variables and macroeconomic variables. All the variables are summarized and explained in Table 1.

In order to examine the factors that influence the profitability of banks in Serbia, Croatia and Slovenia, and factors that influence the profitability of the 10 largest banks in EU, following general model will be used:

$$\pi_{it} = bo_{it} + bm_{it}Ym_{it} + bd_{jt}Yd_{it} + \varepsilon$$
⁽¹⁾

where *i* refers to an individual bank; *t* refers to year; *j* refers to the country in which bank *i* operates; π_{it} the dependent variable that refers to the profitability of the bank (ROAA) and is the observation of a bank *i* in a particular year *t*; *Ym* represents the internal factors; *Yd* represents the external factors and ε is an error term.

Model (1) will be estimated using fixed effects model. The fixed effects estimator considers the within dimension of the data (differences within individuals) while the random effects estimator considers the information from both the between and within dimensions (Verbeek, 2008). Fixed effect model means that the effect of being in a particular group is assumed fixed and represented by a fixed population parameter (Campbell, 2006). Since panel data in this research contains the significant differences among banks with data which may or may not much differ over time, the fixed effect model is the appropriate choice. For testing the appropriateness of fixed effects model instead of random effects model the Hausman test will be used. Hausman test is a statistical hypothesis test in econometrics which evaluates the significance of an estimator versus an alternative estimator. It helps to evaluate if a statistical model corresponds to the data (Wu, 1973).

Using all the variables, model will be formulated as follows:

$$\pi_{it} = bo_t + b_{it}(SIZE_{it} + CIR_{it} + EQAS_{it} + LODEP_{it} + INCEM_{it}) + b_{jt}(ASSGDP_{jt} + MCGDP_{jt} + SPREAD_{jt} + INF_{jt} + GDP_{jt})$$
(2)

Table 1: Description of variables
--

Type of variables	Symbol for the variable	Description
Dependent		
variable	ROAA	Return on average assets
	SIZE	Size of a bank (total assets)
	CIR	Cost to income ratio
Bank-specific variables	EQAS	<i>Capitalization of a bank (equity to total assets)</i>
	LODEP	Liquidity of a bank (loans to deposit ratio)
	INCEM	<i>Productivity of a bank (total income to number of employees)</i>
	ASSGDP	Total assets of the deposit money banks to GDP
Market-specific variables	MCPASS	Market capitalization to total assets of the deposit money banks
	MCGDP	Market capitalization to GDP
	SPREAD	Difference between loan and deposit rate
Macroeconomic	INF	The annual inflation rate
variables	GDP	Gross domestic product growth

4 Conclusion

This study compares the influences of internal and external factors on the profitability of banks in transition economies – Serbia, Croatia, Slovenia and in 10 largest EU banks. Research is based on the observations of individual bank's profitability and it measure influence of 11 different factors.

Since Slovenia has joined EU 8 years ago, this research should show similar results for the Slovenia and for the 10 largest EU banks. Croatia and Serbia are on their way to join EU, and more differences in the factors that affect banks profitability in these countries and largest EU banks are expected.

Furthermore, the contribution of this research is reflected in the fact that the management of banks in Serbia and Croatia will be able to assume which factors will have or will not have an impact on the profitability of their banks in the future.

References

Abreu, M. and Mendes, V. (2002) Commercial bank interest margins and profitability: evidence from E.U countries. *Porto Working paper series*.

Afanasieff, T., Lhacer, P. and Nakane, M. (2002) The determinants of bank interest spreads in Brazil. *Banco Central di Brazil Working Papers*.

Angbazo, L. (1997) Commercial bank net interest margins, default risk, interest-rate risk, and off-balance sheet banking. *Journal of Banking and Finance*. Vol. 21, 55-87 Athanasoglou, P., Delis, M. and Staikouras, C. (2006) Determinations of bank profitability in the South Eastern European Region. MPRA Paper No. 10274 Ben Naceur, S. and Goaied, M. (2001) The determinants of the Tunisian deposit banks' performance. *Applied Financial Economics*. Vol. 11, 317-19

Berger, A. (1995) The Relationship Between Capital and Earnings in Banking. *Journal of Money, Credit and Banking*. Vol. 27, 404-431

Campbell, M.J. (2006) *Statistics at square two*, Second Edition, Blackwell Publishing Inc., USA

Gul, S., Irshad, F. and Zaman, K. (2011) Factors Affecting Bank Profitability in Pakistan. *The Romanian Economic Journal*. Vol. 39, 61-87

Kosak, M. and Cok, M. (2008) Ownership structure and profitability of the banking sector: The evidence from the SEE region. *Zb. rad. Ekon. fak. Rijeka*. Vol. 26, 93-122 Molyneux, P. and Thorton, J. (1992) Determinants of European bank profitability: A note. *Journal of Banking and Finance*. Vol. 16, No. 6, 1173-1178

Neely, M.C. and Wheelock, D. (1997) Why does bank performance vary across states? *Federal Reserve Bank of St. Louis.* Vol. 79, No. 2, 29-40

Pasiouras, F. and Kosmidou, K. (2007) Factors influencing the profitability of domestic and foreign commercial banks in the European Union. *Research in International Business and Finance*. Vol. 21, 222–237

Pejic Bach, M., Posedel, P. and Stajanovic, A. (2009) Profitability determinations of banks in Croatia. *Zbornik Ekonomskog fakulteta u Zagrebu*. Vol. 1, 81-92

Ramlall, I. (2009) Bank-Specific, Industry-Specific and Macroeconomic Determinants of Profitability in Taiwanese Banking System: Under Panel Data Estimation. *International Research Journal of Finance and Economics*. Vol. 34, 160-166

Rumler, F. and Waschiczek, W. (2011) The impact of Economic Factors on Bank Profits. *Monetary Policy & The Economy Q4/2011*. Oesterreichische Nationalbank

Verbeek, M. (2008) *A guide to modern econometrics*, Third edition, John Wiley & Sons Ltd, England

Wu, D.M. (1973) Alternative tests of independence between stochastic regressors and disturbances, *Econometrica*, 41, 733-750

Empirical investigation of credit risk determinants: The case of Tajikistan banking industry

Mukhayo Khodzhibaeva¹, Eftychia Nikolaidou²

¹ Business Administration and Economics Department, CITY College - International Faculty of the University of Sheffield, mkhodzhibaeva@city.academic.gr

² Business Administration and Economics Department, CITY College - International Faculty of the University of Sheffield, enikolaidou@city.academic.gr

Abstract. The purpose of this paper is to empirically investigate the macroeconomic and bank-specific determinants of nonperforming loans (NPL) in the banking industry of Tajikistan for the period from 2000 to 2011. Moreover, recent global financial crisis makes investigation of this topic very interesting as we believe that along with endogenous factors the exogenous one also plays important role in the materializing of credit risk. The authors employ time-series modeling (general-to-specific approach widely known as Hendry's method) with further diagnostic checking of the final model for robustness. The overall conclusion based on empirical findings is that NPLs in Tajikistan are determined by Capital investments, Exchange rate (somoni/USD) and Credit growth; furthermore, the world economic crisis seems to have had a positive influence on the value of NPL.

Keywords: Financial crisis; banking; credit risk; nonperforming loans; Tajikistan

1. Introduction

Typically, the financial system and particularly the banking industry is called the "blood-vascular system" of the economy due to its primary function in mobilizing funds from surplus units and channeling them to the deficit units for productive investments. Therefore, the soundness, efficiency and safety of the financial system are crucial for the health of the economy. This strong connection implies that any problems encountered by the financial institutions can easily become systemic and lead to a crisis that can be long-lasting, hit hard and impact the whole economy.

In this paper we present the empirical investigation of the macroeconomic and bank-specific determinants of non-performing loans in Tajikistan's banking industry over the period 2000-2011. Moreover, the recent global financial crisis makes investigation of this topic very interesting as we believe that along with endogenous factors, the exogenous ones also play important role and might influence credit risk.

The global financial crisis of 2007-2008 caused serious economic damage to most countries in the world. The crisis also brought influence to the economies of Commonwealth of Independent States (CIS) including Tajikistan. However, the impact of the global crisis was varied among countries and mostly it depends on the strength of financial and banking systems, the effectiveness of financial institutions and their external exposure. According to Dadabaeva (2009) the official authorities' position in Tajikistan concerning the financial crisis by the end of 2008 was rather optimistic as they claimed that Tajikistan has its own economic development peculiarities. They asserted that the country does not have direct interaction with the world financial markets; it virtually doesn't have securities market and the production sector is not too significant to be able to damage the economy. However, more precisely, the situation of the financial sector of Tajikistan in the pre-crisis period has been described by the IMF Financial System Stability Assessment report (2008): "the financial system is small and dominated by banks which remain well capitalized and liquid but the rapid growth of their loan portfolios has the negative impact to capital reserves. The interbank money market and the security market are virtually nonexistent. The latest stress test results show a low level of banks resilience. The large stock of distressed cotton debt is still exists". In addition, according to Ratha, Mohapatra and Silwal (2009) the volume of remittances in 2008 makes half of Tajikistan's GDP. Besides, as Dadabaeva (2009) points out the rates of foreign trade turnover worsen due to the reduction of aluminum and cotton exports, the largest source of foreign exchange inflows to the country. Thus, apart from the problems in the economy as a whole, the crisis has also revealed shortcomings in the financial sector.

In the proposed paper we are interested in studying the effects of internal and external driving forces on the level of non-performing loans at the banking system of Tajikistan. Unfortunately, we could not find any thorough academic researches with empirical evidence related to Tajikistan banking sector neither in open access nor as reference to the limited access or at least to the name or articles abstract. Of course, there are available some papers on the topic but mostly with descriptive statistics and mainly related to the economy as a whole without focusing on the banking system. Given the lack of any relevant studies for this country the proposed research aims to fill in the gap in the literature by providing empirical evidence on the issue focusing on Tajikistan. So, the key reasons that justify the selection of this topic can be found in:

- *the topicality of the issue;* credit risk is considered as one of the dominant source of risk for banks and the subject of strict regulatory oversight and policy debate. Nowadays, in the time of continuous financial distresses analysis and investigation of factors determined credit risk is critical for financial sectors of any country because of negative consequences on the economy.

- *the novelty of the topic for Tajikistan;* since available papers mainly do aggregate analysis that covers all countries of Central Asia region and virtually there is no separate study on the banking system of Tajikistan we believe that our research will fill in this gap.

- the in-kind contribution to the development of the country's financial sector; along with investigation of internal determinants of credit risk we also interested in study of external factors and believe that results of this study make a small but significant contribution to the economic development of the country.

The rest of the paper is organized as follows: Section 2 provides literature review on the topic, Section 3 describes the data and methodology used for empirical analysis in the paper, Section 4 demonstrates the process of empirical investigation with interpretation of findings and summary of the whole work is indicated in Section 5.

2. Literature review

The theory suggests that most financial distresses are preceded by activation of various processes in the economy such as the lending boom in the banking industry typically associated with the credit risk which is invisible up to a certain time. The determinants of the credit growth examined in some studies (Saurina and Jimenez, 2006; Festic, Repina and Kavkler, 2009; Erdinç, 2010) are identified as results of classic agency problems (when managers rewards depend on growth objectives), strong competition between banks which lead to deterioration of lending standards, excessive liquidity and lightening of regulatory policy. Thus, take into account such pre-crisis fragility of banks the first impact of the financial distress become apparent through materializing credit risks marked by the increasing volume of nonperforming loans (NPL). At the same way, Gavin and Hausmann (1996) state that along with bank specific factors the macroeconomic forces are also "at least part of the story". Studying the banking crisis roots in Latin America during 1990s the authors find that worsening of the macroeconomic environment affects bank asset quality in a number of ways: through decline in GDP and trade and fluctuation in currency exchange rate.

The financial cataclysms and their harmful consequences for countries showed the necessity of thorough research of risk management and the analysis of credit risk in particular to identify the key vulnerabilities in the financial sector. We can observe that many studies of the last years have a principal tendency to explore the determinants (macro and industry-specific) of credit risk, as one of the main risk in the banking industry, with attempt to assist to policy decision-makers.

Vogiazas and Nikolaidou (2011) researching the topic of post-crisis effects on loan performance, investigating the determinants of credit risk in the Romanian's banking industry. By employing time-series analysis the authors test the influence of macroeconomic-cyclical indicators and bank-specific variables on NPLs level. The results demonstrate that a number of macroeconomic variables such as country's external debt to GDP ratio, M2, inflation, unemployment and contraction and investment expenditure have an impact to credit risk while bank-specific factors along with financial markets and interest rate indicators did not exert any power. Thus, an assumption that reduction in capital influx deteriorates activity in the contraction sector which leads to increased unemployment and further may impact to NPLs was confirmed. The same explanation power of macroeconomic variables (GDP growth and foreign exchange rate) on the bad loans ratio as a measure of credit risk were detected by Cota, Bogdan and Rodic (2011). Researchers investigate the Croatian banking system and validate a supposition of foreign exchange rate influence to assets quality (almost 20% of loans in Croatia are denominated in Swiss Franc). However, analysis of NPL trends of individual countries provide less robust results in comparison with cross-country, argue the authors of ECB Financial Stability Review (2011), because time-series for NPLs commonly span quite short period of time (near ten years) and mostly among emerging countries. So, ECB (2011) based on panel analysis of 80 advanced and emerging economies for the past decade suggests that both rise in real GDP in all countries and exchange rate depreciations in countries with major lending in foreign currency lead to increase NPLs ratio which is consistent with Cota, Bogdan and Rodic (2011) results. In contrast to Vogiazas and Nikolaidou (2011) ECB found that decline in share prices in countries with high and low stock market capitalization tend to negatively affect the banks' asset quality. As for bankspecific indicators, Festic, Repina and Kavkler (2009) testing financial intermediaries in Baltic States, Bulgaria and Romaia prove the positive effect of loans-to-assets ratio to credit quality which implies that the credit boom lead to worsening loans performance.

In turn, Louzis, Vouldis and Metaxas (2010) disclose the influence of other factors to loan performance – the unemployment rate and the quality of management. The interesting outcomes of testing we can observe in the works of Mancka (2012) who proves the hypothesis that the increase of the bad loans are provoked by the high exchange rate while Festic, Repina and Kavkler (2009) obtain opposite results. It is worth stating that many studies (Pesola, 2001; Saurina and Jimenez, 2006; Fainstein and Novikov, 2011) using different methodologies obtained consistent evidence, in particular the impact of the GDP growth is negative and influence of the credit growth is positive with respect to loan losses. On the other hand, Espinoza and Prasad (2010) confirm the supposition of the feedback effect (though short-term) from bank losses on economic activity and this statement is also supported by ECB (2011).

The review of the literature on the topic demonstrates how the authors by mixing various indicators and testing different econometric techniques attempt to "get at the roots of things". Certainly, the findings are valuable and useful for policy and decision makers as point out aspects which require priority attention during management of credit risk.

3. Data and methodology

The general information about the banking system of Tajikistan and the economy as a whole together with knowledge gained from the analysis of the literature on the topic help us in selecting suitable variables and employing the proper methodology.

For the purpose of this paper we use monthly based secondary data for the period 2000-2011 from the official online sources of the National Bank of Tajikistan (NBT) and the Agency on Statistics of Tajikistan.

As empirical literature suggests we select one of the indicators of financial stability, *bank nonperforming loans (NPL)* as the dependent variable. From the theory it well known that NPL is one of the measure of credit risk and, therefore, it will be able to show the efficiency of resource allocation, the quality of loans and the extent of exposure to credit risk. In our case NPLs include uncollateralized loans that are 60 days past due and collateralised ones which are 180 days, debts of clients with unsatisfactory financial situation or that who are declared as bankrupts (NBT, 2011). As similar studies (Gavin and Hausmann, 1996; Erdinç, 2010; Vogiazas and Nikolaidou, 2011) hint it is expected that worsening of the macroeconomic environment as well as rapid growth of loans will increase the volume of NPLs.

Our model includes a number of independent variables represented by several macroeconomic indicators as well as bank-specific one:

Capital investments (CAPINV) are funds invested into productive businesses which drive the economic growth. Deterioration of the investment influx increase probability of loan defaults.

Unemployment (UNEMP) along with bringing economic imbalance to society leaves people unproductive and unable to meet financial obligations. As results, the higher the unemployment the more likely credit risk to increase.

The Industrial production Index (IPI) is selected as proxy for the real economic activity in Tajikistan and IPI's growth is negative with respect to banks' asset quality.

Consumer price index (CPI) as a measure of inflation, allows investors to estimate the total returns to meet their financial goals. Thus, high level of CPI deteriorates the loan portfolio.

Somoni/USD Exchange rate's (EXCHR) adverse changes make contribution to the loan default, in particular if loan disbursed in foreign currencies.

Loan growth (LOANGR) associated with lightening of credit standards; excessive liquidity and relaxing of regulatory policy worsen the quality of credit at banks. Theory says that the loan growth in the past can result into the high level of NPLs in the future. However, Espinoza and Prasad (2010) argue that it also can be the feedback effect when increased level of NPLs forces banking industry to sharply decrease the lending activity which has outcomes to the macroeconomy.

A new indicator variable (CRISIS2007DUM) is a "dummy" variable taking the value of 1 from March 2009 (period of sharp deterioration of socio-economic situation in Tajikistan) are used as proxy of Global financial crisis with expected positive effect on the loan performance.

Taking into account the specificity of the topic we found more appropriate to employ in the empirical part the time-series modelling (general-to-specific approach widely known as Hendry's method where: NPL = f (x1, x2, x3, ...x7)) used by Vogiazas and Nikolaidou (2011) who investigated the impact of the Greek crisis to the Romanian nonperforming loans. According to Brooks (2008) at the starting point a general model should comprises all relevant stationary variables and after testing

those variables that are not statistically significant will be gradually eliminated in order to end up with a model that best explains the dependent variable. Further, based on the statistical significance of coefficients the suitable interpretation will be done. Moreover, depending on the received results and in case of necessity additional diagnostic tests (for heteroscedasticity, autocorrelation and multicollinearity) can be implemented for the robustness of the model. The program used for testing all econometric models in this paper is EView7.

4. Empirical findings

Prior to start the analysis variables (except IPI, CPI and LOANGR which are in %) should be transformed into the logarithm form with the purpose to stabilize the variance in time series. Since Gujarati (2004) warns that regression of non-stationary time series can result to spurious regression problems we use the Augmented Dickey-Fuller (ADF) test to examine variables for existence of a unit root (non-stationarity) in time-series with further conversion those into stationary ones by employing the difference operator. Thus, all variables become stationary after taking the first difference of the series, i.e. variables were integrated of order 1 [~ I (1)], except LOANGR which is stationary at level (see Table 1).

Table 1. Results of the Augmented Dickey-Fuller test (ADF)					
Variablas (in diastors)	Variables are integrated				
Variables (indicators)	of order 0 [~ I(0)]	of order 1 [~ I(1)]			
Non-performing loans		Х			
Capital investment		Х			
Unemployment		Х			
Industrial production Index (IPI)		Х			
Consumer price index (CPI)		Х			
Exchange rate (somoni/USD)	X				
Loan growth	Х				

Following to Vogiazas and Nikolaidou (2011) in order to detect the explanatory power of each independent variable on the NPL we run OLS univariate regression at current time t as well at all lags up to 12 months where needed. Thereby, we attempt to "catch up" possible lagged effect of independent variables on NPLs. In the Table 2 we demonstrate the results of univariate regression with the expected positive (negative) signs suggested by theory.

Table 2. Results of the univariate regressions							
Variables (indicators)	Expected	Reg	gression on lag	gged factor	Xt-s (s=112	2)	
variables (indicators)	sign	Coeff.	t-stat	P-value	R^2	lag(s)	
Capital investment	-	-0.07765	-2.777576	0.0062	0.052584	2	
Unemployment	+	-2.256788	-2.085893	0.0389	0.03191	9	
Industrial production Index (IPI)	-	-0.013641	-1.584959	0.1154	0.018958	10	
Consumer price index (CPI)	-	0.011051	0.597252	0.5513	0.021296	2	
Exchange rate (somoni/USD)	+	3.637466 2.352797 0.0200 0.038036				1	
		Regression on current factor Xt					
		Coeff. t-stat P-value R ² lag					
Loan growth	+/-	0.896979	2.298936	0.0230	0.036129	no lag	

During analysis we identify the statistical significance of coefficients at 5% level by rejecting the hypothesis that coefficient of variable is equal to zero ($\alpha_1=0$). In order to declare that coefficient is statistically significant its t-statistic value should lie in the critical regions (teritical <tstatistic< -teritical), where teritical = (+/-) 1.97. Thus, from the Table 2 we can see that Capital investment at 2nd lag, Unemployment at 9th lag, Exchange rate at 1st lag and Loan growth at current level demonstrate the significance of their coefficients while other variables are insignificant. The p-values which are less than 0.05 also confirm the significance of coefficient and CAPINV can be considered even strongly significant as its P-value is less than 0.01. Checking the consistency of finding with theoretical consideration we notice that three variables CAPINV, EXCHR and LOANGR have the expected directional influence on nonperforming loans except UNEMP. The negative sign of UNEMP can be interpreted as increased level of unemployment decrease the volume of bad loans which make no sense. However, considering the fact that huge stream of unemployed unqualified labor power from Tajikistan go to seasonal work to Russia and then send a large volume of remittance back to Tajikistan (making half of the Tajikistan's GDP) we can at some degree (R^2 -value implies that just 3.19% of the variation in NPL is explained by UNEMP) explain the results of regression but we decide to drop this variable from the final model.

Now, follow to the "driver" (Vogiazas and Nikolaidou, 2011) and using generalto-specific approach we estimate a multivariate model which satisfy the following criteria: (a) the variables in the model have the correct sign; (b) at least one variables from each category (macro, financial market and bank-specific) are included in the model; (c) the model is simple and logically plausible; (d) all variables in the model are significant at the 10% level. In this stage we also add into the model the "dummy" variable which represents the financial crisis of 2007.

Thus, the results of multivariate regression are presented in the Table 3:

Included observations: 141 after adjustments						
Variable	Coefficient	Std. Error	t-Statistic	Prob.		
C DLCAPINV(-2) DLEXCHR(-1) LOANGR CRISIS2007DUM	-0.034167 -0.074244 3.957429 0.922701 0.127969	0.032984 0.026889 1.484330 0.375408 0.076771	-1.035871 -2.761114 2.666138 2.457861 1.666876	0.3021 0.0066 0.0086 0.0152 0.0978		
R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood F-statistic Prob(F-statistic)	0.146648 0.121549 0.323705 14.25071 -38.48765 5.842874 0.000228	Mean dependent var S.D. dependent var Akaike info criterion Schwarz criterion Hannan-Quinn criter. Durbin-Watson stat		0.031057 0.345374 0.616846 0.721412 0.659338 2.023235		

Table 3. Regression output of the multivariate model

Dependent Variable: DLNPL Method: Least Squares Date: 07/04/12 Time: 15:08

Sample (adjusted): 2000M04 2011M12

For the robustness of the model we decided to conduct additional diagnostic tests (see eViews table in Appendix). The results of the Durbin-Watson test and Breusch-Godfrey LM Test indicate no problems with serial correlation. The test for heteroscedasticity doesn't reveal any ARCH effect meaning that there is no heteroscedasticity in the model. Moreover, since the higher correlation coefficient among independent variables is 0.024 which is sufficiently small and can be reasonable ignored (Brooks, 2008) we can state that there is no multicollinearity problem in the model.

Now, let's look at the multivariate regression on Table 3 and interpret results. All three independent variables (DLCAPINV, DLEXCHR, LOANGR and CRISIS2007DUM) are statistically significant as P-value in all cases is less than 0.10. On the other hand, though R^2 is small enough (15%) we cannot say that the model well fit within the sample but at the same time F-statistic (5.842874) with corresponding P-value (0.000228) which is highly significant meaning that the model fits very good within the population.

In this stage we would like more thoroughly look at the global crisis variable as besides showing expected sign it is also statistically significant. It means that our finding empirically prove official reports of NBT and IMF, statements of independent economists that Tajikistan was severely affected by the global crisis starting from the beginning of 2009 mainly through a sharp decline in remittances (1/3) which reduced the real income of the population and weak demand to the main exports of Tajikistan - cotton and aluminum. Moreover, shortage in liquidity along with sharp devaluation of local currency on March 2009 and inability of government's currency intervention

because of too low level of international reserves added stress in the banking sector. The situation was worsening by the harsh energetic crisis (most areas had electricity no more than one to three hours a day). With attempt to archive energy independence the authorities decided to accelerate the construction of the Roghun Hydropower Project by financing this campaign through equity sales to population starting from 2010 which was not welcomed by IMF. As results population and businesses income reduced because near 800 million somoni was withdrawn during campaign and later frozen on the Treasure account. Thus, it seems that deterioration of social-economic situation on that period in Tajikistan confirm the positive impact of the global financial crisis and internal factors to the level of NPL.

Summing up our analysis we conclude that though the independent variables (capital investment lagged by two months, exchange rate lagged by one month together, loan growth together with the global crisis indicator), taken all together, explain just 15% of variability of credit risk in Tajikistan's banking sector during 2000-2011, but based on the highly significant value of F-statistic we can conclude that all the independent variables in the model jointly can influence the dependent variable – Nonperforming loans.

5. Conclusion

The purpose of this paper is to investigate the determinants of nonperforming loans in the banking industry of Tajikistan for the period from 2000 to 2011. After brief presentation of the pre-crisis financial situation of Tajikistan in the introduction part we proceed with the description of data and methodology selected for the purpose of this research. Based on suggestion of empirical literature we selected several macroeconomic and bank-specific indicators with the purpose to identify their influence on the quality of banks assets for the given period. This study also motivated by the supposition that the recent global financial crisis may have positive influence to the decreased level of non-performing loans. Taking into account the specificity of the topic we found more appropriate to employ general-to-specific approach widely known as Hendry's method, used by Vogiazas and Nikolaidou (2011). Thus, during the empirical analysis we replicated all steps implemented by authors. After checking the variables for presence of a unit root we transformed them where needed into stationary ones. In order to identify the significance of each independent variable we run the univariate regression of the regressors' current value and where necessary their lagged values against NPL variable. Then we estimated multivariate regression with those significant variables and the dummy variable which proxies the global financial crisis. However, we decided not to include the unemployment variable into the final model as it does not show expected sign which is contradict to theoretical expectation. As a result we ended up with a model which satisfied to the requirement of general-to-specific approach. Additional diagnostic tests did not detect any problems with autocorrelation, heteroscedasticity, and multicollinearity of the model. Thus, the overall conclusion based on empirical findings is that though the capital investments lagged by two months and exchange rate lagged by one month, loan growth together with financial crisis variable, taken simultaneously explain only 15% of variability of credit risk in Tajikistan's banking

sector during 2000-2011, the highly significant value of F-statistic suggests that all the independent variables in the model jointly may affect nonperforming loans.

References

Brooks, C. (2008) *Introductory Econometrics for Finance*, 2nd ed., Cambridge University Press, New York.

Cota, B., Bogdan, Z. and Rodic, L. (2011) Macroeconomic stress testing of Croatian banking system. *Proceedings of the Ninth International Conference Challenges of Europe: Growth and Competitiveness – Reversing the Trends"*, Split: Faculty of Economics in Split, p.133-156.

Dadabaeva, Z. (2009) The impact of the global financial crisis to the economy of Tajikistan. The global financial crisis in the post-soviet countries: national specifics and economic consequences. *Collection of articles. Institute of Economics. Russian Academy of Science*, p. 227-255.

Ratha, D., Mohapatra, S. and Silwal, A. (2009) Migration and remittance trends 2009. A better-than-expected outcome so far, but significant risks ahead. *Migration and development brief*, (11). World Bank

Erdinç, D. (2010) Does a Credit Boom Increase Bank Fragility? Transitional Challenges in Bulgarian Banking, 1999–2008. *Eastern European Economics*, 48 (3), p. 68–87

Espinoza, R. and Prasad, A. (2010) Nonperforming Loans in the GCC Banking System and their Macroeconomic Effects. Working Paper WP/10/224, IMF

European Central Bank (2011) Empirical determinants of non-performing loans. *Financial Stability Review*, Dec. 2011

Fainstein, G. and Novikov, I. (2011) The Comparative Analysis of Credit Risk Determinants in the Banking Sector of the Baltic States. *Review of Economics & Finance*. Article ID: 1923-7529-2011-03-20-26.

Festic, M., Repina, S. and Kavkler, A. (2009) The overheating of five EU new member states and cyclicality of systemic risk in the banking sector. *Journal of Business Economics and Management*, 10(3), p. 219-232

Gavin, M. and Hausmann, R. (1996) *The Roots of Banking Crises: The Macroeconomic Context.* Working Paper 318, Inter-American Development Bank

Gujarati, D. (2004) Basic Econometrics, 4th ed, McGraw-Hill.

International Monetary Fund (2008) Republic of Tajikistan: Financial System Stability Assessment. *IMF Country Report No. 08/371*

Louzis, D., Vouldis, A. and Metaxas, V. (2010) *Macroeconomic and bank-specific determinants of non-performing loans in Greece: a comparative study of mortgage, business and consumer loan portfolios.* Working paper N 118, Bank of Greece

Mancka, A. (2012) The Impact of National Currency Instability and the World Financial Crisis in the Credit Risk. The Case of Albania. *Journal of Knowledge Management, Economics and Information Technology*, (8), pp. 1-18

National Bank of Tajikistan (2011) Instruction #177 'Procedure of creating and using provisioning (Reserve) and loan loss provision fund', NBT

Pesola, J. (2001) *The role of macroeconomic shocks in banking crises*. Bank of Finland Discussion Papers, June 2001

Saurina, J. and Jimenez, G. (2006) *Credit Cycles, Credit Risk, and Prudential Regulation.* MPRA Paper No. 718, University Library of Munich, Germany

Vogiazas, S. and Nikolaidou, E. (2011) Investigating the Determinants of Nonperforming Loans in the Romanian Banking System: An Empirical Study with Reference to the Greek Crisis. *Economics Research International*, Vol. 2011, Article ID 214689

Appendix

1. Breusch-Godfrey Serial Correlation LM Test (1 lag):

F-statistic	Prob. F(1,135)	0.8589
Obs*R-squared	Prob. Chi-Square(1)	0.8556
Test Equation:		

Dependent Variable: RESID Method: Least Squares Date: 07/06/12 Time: 11:40 Sample: 2000M04 2011M12 Included observations: 141 Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C DLCAPINV(-2)	-3.40E-05 6.79E-06	0.033103	-0.001027	0.9992
DLEXCHR(-1)	0.023067	1.495259	0.015426	0.9877
LOANGR CRISIS2007DUM	-0.003982 -0.000184	0.377414 0.077053	-0.010551 -0.002388	0.9916 0.9981
RESID(-1)	-0.015475	0.086870	-0.178143	0.8589
R-squared	0.000235	Mean dependent var		5.51E-18
Adjusted R-squared	-0.036793	S.D. dependent var		0.319047
S.E. of regression	0.324863	Akaike info criterion		0.630795
Sum squared resid 14.247		Schwarz criteri	on	0.756275
Log likelihood -38.47108		Hannan-Quinn criter.		0.681786
F-statistic	0.006347	Durbin-Watson	stat	1.993289
Prob(F-statistic)	0.999990			

2. Breusch-Godfrey Serial Correlation LM Test (2 lag):

F-statistic	0.015750	Prob. F(2,134)	0.9844
Obs*R-squared	0.033138	Prob. Chi-Square(2)	0.9836
Test Equation: Dependent Variable: RESID Method: Least Squares Date: 07/06/12 Time: 11:4' Sample: 2000M04 2011M12 Included observations: 141 Presample missing value lag	1	ls set to zero.	

Variable	Variable Coefficient		t-Statistic	Prob.
C DLCAPINV(-2)	-3.48E-05 6.48E-06	0.033248 0.027090	-0.001045 0.000239	0.9992 0.9998
DLEXCHR(-1) LOANGR	0.023136 -0.003970	1.505058 0.379340	0.015372 -0.010465	0.9878 0.9917
CRISIS2007DUM RESID(-1)	-0.000183 -0.015476	0.077345	-0.002372	0.9981
RESID(-2)	-5.36E-05	0.087074 Mean depende	-0.000616	0.9995
Adjusted R-squared S.E. of regression	-0.044531 0.326073	S.D. dependent var Akaike info criterion		0.319047
Sum squared resid Log likelihood	14.24736 -38.47108	Schwarz criterion Hannan-Quinn criter.		0.791372 0.704469
F-statistic Prob(F-statistic)	0.005250 0.999999	Durbin-Watson	******	1.993296

3. Heteroskedasticity Test: ARCH (1 lag)

F-statistic	Prob. F(1,138)	0.7965
<u>Obs</u> *R-squared	Prob. Chi-Square(1)	0.7947

Test Equation: Dependent Variable: RESID*2 Method: Least Squares Date: 07/06/12 Time: 11:43 Sample (adjusted): 2000M05 2011M12 Included observations: 140 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C RESID^2(-1)	0.104010 -0.021992	0.048927 0.085095	2.125811 -0.258436	0.0353 0.7965
R-squared	0.000484	4 Mean dependent yar		0.101788
Adjusted R-squared	-0.006759			0.567986
S.E. of regression	0.569902	2 Akaike info criterion		1.727478
Sum squared resid	44.82080	Schwarz criteri	on	1.769502
Log likelihood	-118.9235	Hannan-Quinn	criter.	1.744555
F-statistic	0.066789	Durbin-Watson	stat	2.001183
Prob(F-statistic)	0.796456			

4. Heteroskedasticity Test: ARCH (2 lag)

F-statistic		Prob. F(2,136)	0.9365
Qbs*R-squared		Prob. Chi-Square(2)	0.9352
Que R-squareu	0.134070	FIDD. CHI-Square(2)	0.9352

Test Equation: Dependent Variable: RESID^2 Method: Least Squares Date: 07/06/12 Time: 11:44 Sample (adjusted): 2000M06 2011M12 Included observations: 139 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C RESID^2(-1) RESID^2(-2)	0.106982 -0.022687 -0.021711	0.050249 0.085720 0.085720	2.129046 -0.264658 -0.253279	0.0351 0.7917 0.8004
R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood F-statistic Prob(F-statistic)	0.000965 -0.013727 0.573882 44.79036 -118.5250 0.065654 0.936484	Mean depende S.D. dependen Akaike info crit Schwarz criteri Hannan-Quinn Durbin-Watson	t var erion on criter.	0.102463 0.569984 1.748562 1.811896 1.774299 1.999292

5. Multicollinearity test

Correlation				
	DLCAPINV(-2)	DLEXCHR(-1)	LOANGR	CRISIS2007DUM
DLCAPINV(-2)	1.000000	0.024321	-0.057218	-0.030288
DLEXCHR(-1)	0.024321	1.000000	-0.048784	0.016123
LOANGR	-0.057218	-0.048784	1.000000	-0.062473
CRISIS2007DUM	-0.030288	0.016123	-0.062473	1.000000

US Corporate Bond Yield Spread: A default risk debate

Noaman Shah¹

¹Laboratoire d'Economie d'Orléans (Faculté de Droit d'Economie et de Gestion), Université d'Orléans Email: syed-muhammad-noaman.shah@etu.univ-orleans.fr

Abstract. According to theoretical models of valuing risky corporate securities, risk of default is primary component in overall yield spread. However, sizable empirical literature considers it otherwise by giving more importance to non-default risk factors. Current study empirically attempts to provide relative solution to this conundrum by presuming that problem lies in the subjective empirical treatment of default risk. By using post-hoc estimator approach of Lubotsky & Wittenberg (2006), we construct an efficient indicator for risk of default, by using sample of 252 US non-financial corporate data (2000-2010). On average, our results validate that almost 48% of change in yield spread is explained by default risk especially in recent financial crisis period (2007-2009). Hence, our results relatively suggest that potential problem lies in the ad-hoc measurement methods used in existing empirical literature.

Keywords: *Default risk, credit spread, risk-aversion, measurement error, index construction.*

1 Introduction

Recent financial crisis has led regulators to increase their attention towards controlling credit risk² component of overall corporate structure of financial and non-financial institutions.³ In addition, yield spread provides support in evaluating whether potential and existing investors are adequately compensated for the risk they are bearing or not, particularly in period of poorer economic outlook which in turn

¹Laboratoire d'Economie d'Orléans UMR6221 Faculté de Droit d'Economie et de Gestion, Rue de Blois – BP 26739, 45067 ORLEANS Cedex2, tél : (33) (0)2 38 41 70 37, fax : (33) (0)2 38 41 73 80.

²We will use terms of credit risk, default risk and risk of default interchangeably in this paper.

³Schuermann, T 2004, 'What do we know about loss given default', in *Credit Risk: Models and Management,* Shimko, D, Incisive Media, London.

correlated with likelihood of lower-wealth situations invigorating reduce willingness to bear and hold risk by risk-averse investors in capital market.

This aporia lead theoretical financial economists⁴ to devise efficient methods to adequately value risky nature of corporate securities. The way theoretical models price these corporate debt securities depend fundamentally upon their related credit risk. On the contrary, extensive empirical literature⁵ on the issue finds it practically arduous to accept this conjecture. The purpose of this paper is to attempt relatively to fill the gap by conjecturing that problem lies in adhoc measurement techniques employed by existing empirical literature.

The main attribute of current empirical study is to efficiently minimize measurement errors caused due to ad-hoc treatment of default risk in existing empirical literature. By objectively considering risk of default as linear combination of significant proxy factors, current work through post-hoc estimator approach of Lubotsky and Wittenberg (2006) relatively provide solution to this problem.

The rest of paper is organized as follows. Next section delineates briefly related literature and formulates default risk proxy factors. Section 3 outlines estimation methodology and data used. Empirical result analysis is presented in Section 4. Finally, Section 5 sums up the main findings and concludes current work.

2 Literature Review

Extensive literature exists on the issue of explaining probable components of corporate bond yield spread. Historically, related literature divides them into two main categories; risk of default and residual risk premium attributing to non-default factors (Fisher 1959). But with the advent of option pricing method of Black and Scholes (1973), new horizon emerges to value risky corporate liabilities. Following Black and Scholes (1973), Merton (1974) gives contingent claim theory by evaluating credit risk of corporate debt as a function of call option on its equity securities. In his seminal work on pricing of risky corporate debt, Merton (1974) mainly focuses to incorporate corporate debt valuation process on the basis of risk of default and non-stochastic interest rate. Later, Black and Cox (1976), Longstaff and Schwartz (1995), among others,⁶ also treat risk of default as their primary pillar in deriving new valuation models incorporating and relaxing assumptions⁷ made by Merton (1974) to value corporate securities.

Despite immense efforts made by theoretical authors in treating risk of default as primary focus for valuing risky corporate securities, most empirical work do not confirm their premise to give key weight to default risk as the major component of

⁴Black and Scholes (1973), Merton (1974), Black and Cox (1976), among others.

⁵Elton et al (2001), Collin-Dufresne et al (2001), Delianedis and Geske (2001), Huang & Huang (2003), Tsuji (2005), among others.

⁶Chance (1990), Duffie (1999), Leland and Toft (1996), Ericsson and Renault (2006), Liu et al (2009), Longstaff, Mithal and Neis (2005), Couderc, Renault, Scaillet (2007).

⁷Merton (1974) assumes firm will default only when they fully exhaust all their asset value and include constant interest rate in his valuation model.

yield spread. Elton et al (2001), a pioneer empirical work in disentangling the determinants of corporate yield spread show that expected default loss (risk of default proxy) does not contribute to overall yield spread more than 25 percent. Among many others, Jones, Mason and Rosenfeld (1984), Collin-Dufresne et al (2001), Delianedis and Geske (2001), Huang and Huang (2003), Tsuji (2005), Liu et al (2009), also show that, in general, risk of default does not contribute majorly in explaining overall yield spread as compared to other non-default factors.

Lack of consensus, between theoretical and empirical proponents, in treating risk of default as a major determinant in yield spread spur the question: Why risk of default is not being proven empirically as a major component in explaining yield spread?.

We argue, in this paper, that one of the main concerns for investor is the risky nature of security for which they ought to be satisfied through the inclusion of default risk premium with other non-default related return, i.e. 'greater the security's risk of default, the greater its default premium should be', Sharpe, Alexander and Bailey (1999: 447). Hence, we conjecture that this controversial issue of empirically lacking to accept risk of default as a major proponent of yield spread is mainly due to the measurement error induced by the usage of different nature of ad hoc proxies in existing empirical literature. Therefore, we hypothesize that investor's perception of risk of default is a function of more than ad-hoc proxies utilized in previous empirical work.

To serve this purpose, we treat yield spread as a linear function of optimal explicit factors of default risk and control for the factors associated with overall macroeconomic and financial change. In other words, we focus on to build an optimal index for default risk on the basis of these proxies. This in turn enables to assist in solving the subjective treatment of default risk in overall US corporate bond yield spread.

2.1 Default Risk Proxies

By convention there is no standard definition of what constitutes a risk of default.⁸ Most of the time in existing literature risk of default is being treated on ad-hoc basis. In order to streamline the optimal effect of default risk, we follow the Bank for International Settlements definition of default event (BIS, 2001: 30).⁹ In particular, what we are disposing here is that in order to simplify the conundrum of subjective treatment of default risk, we should treat risk of default as an optimal composite indicator of underlying explicit proxies, which leads to show relative true significance of default risk in overall formulation of yield spread. Bearing in mind the above discussion, we delineate explicit proxy factors of default risk (derived from BIS definition) alongside the control variables.

⁸Schuermann, T 2004, 'What do we know about loss given default', in *Credit Risk: Models and Management,* Shimko, D, Incisive Media, London.

⁹Basel Committee on Banking Supervision 2001, 'The Internal Ratings Based Approach', (Available at http://www.bis.org/publ/bcbsca05.pdf).

2.1.1 Change in Distance to Default (ΔDD)

This proxy takes into consideration net worth of US corporates and volatility in its stock price. We calculate distance to default as a difference between natural logarithm of total assets and total liabilities divided by its monthly stock price volatility. It, ideally, represents the percentage change in overall long term default probability of the debt issuer (i.e. inability to fulfill its long term obligations). We expect an inverse relation between distance to default and the overall yield spread.

2.1.2 Change in Z-score (ΔZ)

Z-score approach predicts corporate probability of default on short term basis by using firm-specific performance ratios and plausibly caters financial health of the obligor past due 90 days or more on any of its credit obligation (Altman 1968). We presume that a positive percentage increase in Z-score expects to lower the corporate bond yield spread.

2.1.3 Change in Credit ratings (Δ IR)

In similar vein, credit rating variations help to cater investors' immediate decision in investing and financing activities. We include credit rating proxy because it presents rating agencies' appraisal of risk of default associated with related corporate fixed income security issues. To represent credit rating changes, we use change in downgrade credit rating percentage because it is suitable to assess the impact of credit rating change on overall risk of default. In addition, increase in downgrade credit rating change will certainly impact positively on the overall change in yield spread for corporates'.

2.1.4 Cash Flow volatility (CFv)

On the other hand to evaluate firms' position from cash flow perspective, we include fluctuations in operational cash flow situation of debt issuer to assess their ability to service outstanding debt as and when it becomes due. We expect a positive relationship between firm cash flow volatility and the risk-averse investor expected return from the risky corporate security. Following, Tang and Yan (2010), cash flow volatility is calculated as the coefficient of variation of operating cash flows by dividing monthly standard deviation to its absolute mean. In particular, we expect that this proxy will cater the true effect of firm's ability to meet its financial obligation in real terms.

2.1.5 Control variables

In order to evaluate the true effect of change in default risk proxy factors on change in yield spread, we control for different proxy variables reflecting the US macroeconomic and financial market condition. Following Tsuji (2005), as volatility in economic activity leads to affect investors' overall perception of required risk

premium, we use change in Industrial Production Index rate (Δ IPI)¹⁰ as a control proxy for change in business cycle (Collin-Dufresne et al 2001). Furthermore, to control for the effects of inflation level, we use change in consumer Price Index (Δ CPI).¹¹ We expect a negative impact of change in inflation level on the explained variable (Merton 1974). This negative relationship is also confirmed by Longstaff & Schwartz (1995) and Collin-Dufresne et al. (2001).

On the other hand, to control for financial market indicators, we use change in fed fund rate to gauge the inter-bank market effect. In addition, we control for last month change in yield spread on its current month value following Van Landschoot (2004). We expect a positive relation between these values.

3 Data and Methodology

3.1 Data

In order to measure optimal effect of risk of default on overall yield spread, monthly data on variable of interest (i.e. default risk proxies) and control variables, outlined in previous section, has been used for the period of 2000 to 2010. This period is suitable for the study as it cited occurrence of two recent financial crises (dot com and subprime crises) in the US economy. For control variables and corporate bond yield spread, entire data is collected from online source of Federal Reserve Bank of St. Louis. On the other hand, for default risk proxy indicators we use Thomson Reuter database to calculate distance to default, cash flow volatility and z-score, individually, for 252 US non-financial corporates. In particular, using New York Stock Exchange (NYSE) Bond Master service we came to include the sample of 252 non-financial US corporates listed with plain vanilla bonds of medium and long term maturities (greater than one year till twenty years). In addition, downgrade percentage change in their credit ratings data is being collected from Standard & Poor's research report 2011.

Further, to collect stock prices (daily and monthly) for US corporate's in order to calculate default risk proxy factors, we use Yahoo! Finance website. Additionally, in order to obtain average monthly time series of these explicit proxy factors we include simple weighted average across individual firms'.

In addition, downgrade credit ratings data is available in annual frequency with US GDP data in quarterly frequency. We use cubic spline method to convert data into monthly frequency in order to make these time series in harmony with the rest of data variables. Since, our data follow time series we checked whether it has a unit root or not. Following Phillips-Perron test (1988)¹², it is observed that almost all of the time series were stationary integrated with zero order¹³. In order to condense the effect of

¹⁰We also use change in GDP (Δ GDP), instead of change in industrial production index (Δ IPI) in order to evaluate robustness and statistical significance of default risk factors.

¹¹We also use change in producer price index (Δ PPI) instead of change in consumer price index (Δ CPI) for robustness.

¹²We follow Phillips-Perron test of unit root.

¹³Because our objective is to analyze relative change in explanatory variables on dependent variable.

cubic spline algorithm we, in our preliminary model, use natural logarithmic form of change in downgrade credit ratings variable and natural logarithmic form of change in GDP control variable. Further, for monthly yield spread data, we use US treasury monthly yield data (averaged on 1, 2, 5, 10 and 20 years) and US corporate bond monthly yield data (averaged on investment and non-investment grades of similar maturity) from the online source of Federal Reserve Bank of St. Louis. Finally, this gives us total of 132 monthly observations of the focused time series data.

3.2 Methodology

Following our premise of objective treatment of default risk, we follow "post-hoc" estimator approach of Lubotsky and Wittenberg (2006). Using multiple regressions (least square) of explanatory proxy variables on dependent variable, Lubotsky and Wittenberg (2006) use regression coefficients of each explanatory proxy to compute a weighted average index, showing true effects of variable of interest. By assuming risk of default as our focused variable, we aim to evaluate the optimal effect of explicit default risk proxy factors on change in the relative yield spread while controlling for macroeconomic and financial market variables.

Since, weights used here are ratio of covariances of dependent variable (Δ YS) and statistically significant default risk proxy variables to the covariance between the most significant proxy variable¹⁴ and dependent variable, this technique helps in optimally minimizing variance of errors terms of these proxy variables that leads to construct our required composite indicator. Thus, it enables us to reduce noise or measurement error caused, in the existing empirical literature, due to using biased coefficients of ad-hoc proxies treated as the risk of default.

Hence, we use this statistic as an optimal indicator reflecting the true effect of risk of default by utilizing previously explained proxies simultaneously through multiple regressions. Therefore, we first evaluate, individually, effects of our default risk proxy factors on change in yield spread and then following our principal objective, evaluate linearly effect of our default risk proxy factors, aggregately, on change in overall yield spread.

This can be manifested as;

$$\Delta YS_t = \theta + \alpha \Delta X_t + \gamma \Delta W_t + \xi_t$$

Whereas, θ represents constant term, ΔX_t represents matrix (vector) of monthly change in default risk factor time series (t-1 to t), ΔW_t represents matrix (vector) of change in control variable time series from month t-1 to month t, α is a coefficient matrix (scalar) of our explanatory proxy variables, γ is a coefficient matrix (scalar) of control variables, ξ_t shows error term following i.i.d, and ΔYS_t represents change in overall yield spread (i.e. difference between US corporate bond yield—non financial and US treasury yield, on average, month t-1 to month t of similar maturity.

(1)

¹⁴We follow Akaike information criteria (AIC) to select the principal proxy factor.

We can describe the final form of our model as;15

$$\Delta YS_{t} = \theta + \alpha_{1}CFv_{t} + \alpha_{2}\Delta DD_{t} + \alpha_{3}\Delta Z_{t} + \alpha_{4}\Delta lR_{t} + \gamma_{1}\Delta CPI_{t} + \gamma_{2}\Delta IPI_{t} + \gamma_{3}\Delta FFR_{t} + \gamma_{4}\Delta YS_{t-1} + \xi_{t}$$
⁽²⁾

Whereas, θ represents constant term, Δ shows monthly percentage change from month t-1 to month t, α represents coefficients of explanatory proxy variables, γ shows control variables coefficients, ξ is a disturbance term following i.i.d and ΔYS_t represents change in overall yield spread.

Following, Lubotsky and Wittenberg (2006), we treat risk of default as a composite explanatory variable¹⁶ which is not been optimally measured in existing empirical literature. Thus it is being represented by an optimal mix of significant proxy factors. Stated differently, to show the optimal effect of risk of default factor on change in yield spread of non-financial US corporate securities, we use the post-hoc estimator approach.

In addition, our presumption delineates the objective treatment of risk of default based on proxy factors derived from BIS definition of default event. On the basis of equation (2), we include those proxy variables which are statistically significant in order to construct our representative indicator for risk of default.

$$\rho_{t} = \sum_{j=1}^{k} \frac{Cov(Y_{t}, X_{jt})}{Cov(Y_{t}, X_{1t})} * \beta_{j}$$
(3)¹⁷

In equation (3), ρ_t shows our post-hoc estimator for the risk of default indicator in month *t*. Further, Y_t represents our dependent variable (change in yield spread— Δ YS) in month *t*, and X_{jt} represents our statistically significant proxy factors *j* in month *t*, whereas β_j represents the least square regression coefficient of significant *j* explanatory variables at time *t* (months). Further X_{lt} presents principal statistically significant explanatory variable¹⁸ on which the post-hoc estimator is constructed.

Equation (3) gives us the optimal proxy indicator for default risk by minimizing measurement errors in order to evaluate its true effect on the overall yield spread formulation for risky corporate securities. To assess this effect, we regress US corporate bond yield spread on our post-hoc risk of default estimator.

$$\Delta YS_t = \phi + \lambda \Delta \rho_t + \upsilon_t$$

In equation (4), YS_t represents difference between US corporate bond yield (investment and non-investment grade) and the US treasury yield, on average, in month *t*, ρ_t represents the default risk post-hoc estimator in month *t*, whereas λ shows

(4)

¹⁵Following Collin-Dufresne et al (2001) empirical model.

¹⁶" X_t " in equation (1) represents risk of default indicator.

¹⁷Detail derivation can be found in Lubotsky, D and Wittenberg, M 2006, 'Interpretation of Regressions with Multiple Proxies', *The Review of Economics and Statistics*, vol. 88, no. 3, pp. 549-562.

¹⁸We follow Akaike information criteria (AIC) to select the principal factor as X_{1t}.

the regression coefficient of our default risk estimator and \mathcal{U}_t represents i.i.d error term. \varDelta represents the monthly percentage change in variables (from month t-1 to month t). Further, ϕ shows the constant term in equation (4). In order to evaluate robustness of our default risk optimal indicator, we take into consideration its effect on overall yield spread during recent US financial crisis (July 2007 to March 2009). We divide our sample time interval into crisis and non-crisis period and analyze our explanatory variable effect on dependent variable.¹⁹ For this we introduce dummy variable into specification (4) and our estimated model becomes;

$$\Delta YS_t = \phi + \lambda_1 \Delta \rho_t * D_{crisis} + \lambda_2 \Delta \rho_t * D_{non-crisis} + \upsilon_t$$
(5)

In equation (5), ϕ shows constant term, Δ represents monthly percentage change in variables (from month t-1 to month t), YS_t is the yield spread, on average, in month t, ρ_t represents the default risk post-hoc estimator in month t, D_{crisis} is a dummy variable taking value of one for the months during financial crisis, while $D_{\text{non-crisis}}$ is a dummy variable taking value of one for months in normal (i.e. non-crisis) period. v_t shows disturbance term with i.i.d.

Specification (5) mainly explains how change in our risk of default indicator affects change in overall yield spread during crisis and non-crisis period. We use interaction term of our focused estimator with the crisis and non-crisis dummy variable in order to effectively gauge the behavior of our explanatory variable on dependent variable during the crisis period. Next section delineates and analyzes the econometric results and limitations of specifications introduced in this section.

4 Empirical Analysis

Our preliminary objective is to study monthly percentage change of the default risk explanatory proxies on our dependent variable i.e. change in yield spread, therefore, focused time series become stationary integrated with order zero after taking percentage change effect.²⁰

Since our explanatory proxy variable time series are stationary, we follow ordinary least square (OLS) regression for the econometric estimation. Table (1), column (1) shows the result of introducing cash flow volatility as a risk of default proxy to explain changes in yield spread (while controlling for macroeconomic and financial market factors). In column (1), the operational cash flow volatility of non-financial US firms significantly affect change in relative yield spread following presumed relation as delineates in section 2.1.4.

In particular, if we glances overall results in all columns of table (1), then it is evident that all our explanatory proxy variables follow expected relationship with our dependent variable. Furthermore, in column (2), (3) and (4) we evaluate, individually,

¹⁹We deeply thank anonymous reporter comments regarding taking into consideration recent US financial crisis disruptions in our analysis.

 $^{^{20}}$ We use Augmented Dickey Fuller and Philips Perron tests of stationarity to check the presence of unit roots. All our time series are stationary integrated with order zero i.e. I (0).

the effects of change in distance to default (Δ DD), change in Z-score (Δ Z) and change in downward grade credit rating (Δ IR) on our explained variable (i.e. Δ YS). We find that Δ DD and Δ IR significantly affects Δ YS. In addition, Δ Z although shows expected relation with dependent variable is not significant. On the other hand, if we look at our control variables in column (1) to (4), we find that change in inter-bank interest rate and change in previous period yield spread (month t-1) significantly affects the change in current yield spread (month t).

$\begin{array}{c c c c c c c c c c c c c c c c c c c $	yield spread change						
$\begin{array}{cccccccccccccccccccccccccccccccccccc$		1	2	3	4	5	6
$\begin{array}{cccccccccccccccccccccccccccccccccccc$							
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	CFv						1.30 ^c
$\begin{array}{cccccccccccccccccccccccccccccccccccc$		(0.37)				(0.92)	(0.79)
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	ΔDD		-0.15 ^a			-0.15 ^a	-0.17^{a}
$\begin{array}{cccccccccccccccccccccccccccccccccccc$			(0.03)			(0.03)	(0.03)
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	ΔZ			-0.07		-0.04	-0.02
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$				(0.05)		(0.03)	(0.03)
Control Δ CPI -0.15 -1.97 -0.63 -1.89 -4.65 ^b (2.02) (1.92) (1.99) (2.21) (2.09) Δ IPI -0.09 -0.43 -0.31 -0.01 -0.65 (1.02) (0.95) (1.01) (1.00) (0.92) ΔFFR -3.24 ^a -2.55 ^a -3.18 ^a -3.22 ^a -2.54 ^a -2. (0.56) (0.54) (0.55) (0.52) $(1.$ -1. ΔIGDP - - - -1. (0. ΔSt-1 1.47 ^a 1.24 ^a 1.62 ^a 1.57 ^a 1.34 ^a 1.2. Δ1GDP - - - - - - - ΔYSt-1 1.47 ^a 1.24 ^a 1.62 ^a	ΔlR				0.90^{b}	1.17^{a}	1.21 ^a
$\begin{array}{cccccccccccccccccccccccccccccccccccc$					(0.45)	(0.42)	(0.49)
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Control				()		· /
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\Delta \text{ CPI}$	-0.15	-1.97	-0.63	-1.89	-4.65 ^b	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$		(2.02)	(1.92)	(1.99)	(2.21)		
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Δ IPI	· /	· · · ·	· · ·	· · · ·		
$\begin{array}{cccccccccccccccccccccccccccccccccccc$		(1.02)	(0.95)	(1.01)	(1.00)		
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	ΔFFR						-2.62 ^b
$\begin{array}{cccccccccccccccccccccccccccccccccccc$							(1.56)
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	ΔΡΡΙ	(****)	(*****)	(****)	(0.000)	(***=)	-1.82^{a}
$\begin{array}{cccccccccccccccccccccccccccccccccccc$							(0.61)
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	AIGDP						-1.79
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$							(1.2)
(0.55) (0.51) (0.54) (0.5) (0.5) Observations 131	AYS: 1	1.47^{a}	1.24^{a}	1.62^{a}	1 57 ^a	1 34 ^a	1.25 ^b
Observations 131 <t< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td>(0.53)</td></t<>							(0.53)
R² Adjusted 0.22 0.2 0.15 0.24 0.37 0 D-W 1.79 1.91 1.88 1.83 1.99 1. Breusch-Godfrey LM Test 1 <td>Observations</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>131</td>	Observations						131
D-W 1.79 1.91 1.88 1.83 1.99 1. Breusch-Godfrey LM Test			-	-		-	0.3
Breusch-Godfrey LM Test							1.97
LM Test		1.17	1,71	1.00	1.00	1.77	1.77
	5						
						15 50	16.97
							0.15
Note: Standard errors are reported in brackets. Column (1) to (4) presents individ			ntad in hear	Irota Cal-	(1) to (4)		

Table 1: Influence of change in risk of default proxy factors on US Corporate bond yield spread change (ΔYS_t)

Note: Standard errors are reported in brackets. Column (1) to (4) presents individual explanatory variable change effect on the change in dependent variable. Column (5) reports aggregate explanatory variable change effect on the dependent variable change whereas column (6) results are used for robustness with different control variables. Whereas a, b & c represent significance level at 1%, 5% & 10% respectively.

Table (1), column (5), report results of our specification (2), which delineates the linear effect of change in default risk proxy factors on change in overall yield spread, while catering the effects of control variables. After reviewing the results of first four columns of table (1), we may be factual to validate our conjecture of treating risk of default as a linear combination of these proxy factors as their explanatory power significantly enhances to 37 percent in rationalizing overall variation in US corporate bond yield spreads.

Following the first four columns, we evaluate the results of our main model (column 5, table 1). In column (5), we discover that cash flow volatility (CFv), change in distance to default (Δ DD) and change in downgrade credit rating (Δ IR) remains statistically significant. On the contrary, change in Z-score (Δ Z) although reports presumed relation with change in yield spread (Δ YS) remains statistically insignificant.

Thus, results in column (5) table (1), describe significant long term effect of default risk proxy factors on the dependent variable. In general, one percent increase in distance to default for US non financial firms (i.e. showing that a firm is moving away from the level of default) reduces related yield spread percentage by almost 0.153 units. Further, a one percentage point increase in operational cash flow volatility increases vulnerability of technical insolvency of US non-financial firms and hence raises their respective yield spread by almost 1.51 units. In similar vein, a percentage increase in downgrade credit rating increases the overall yield spread by almost 1.2 units, validating presence of risk-averse investors in capital market.

On the other hand, with the linear introduction of default risk proxy factors, we find that change in inflation (Δ CPI) becomes statistically significant. Whereas, change in interest rate (Δ FFR) effect and change in last month yield spread (Δ YS_{t-1}) remains statistically significant, respectively.²¹ In particular, our results confirm the relationship between inflation and interest rate on overall yield spread put forward by theoretical strand o f related literature. Furthermore, Breusch-Godfrey Lagrange Multiplier test (B-G test) suggests that we cannot reject null hypothesis of no serial correlation in residual error terms (i.e. p-value of 0.215).

Column (6), table (1), for robustness, show the results which are not significantly different from results of column (5).²² In addition, B-G test for column (6) also reports no evidence of significant serial correlation in our residual error terms.

On the other hand, we did not find any significant correlation between default risk proxy factors that indicate presence of any multicollinearity problem. Furthermore, correlations between these proxy variables follow, in general, their expected signs. But in order to be certain that results reported in table (1) are not exacerbated by this

²¹As constant term (θ) is statistically insignificant with minimal regression coefficient, we drop related results.

²²We replace control variables, further, we also introduce, simultaneously, change in stock market return (Δ SP), S&P500, and change in its volatility index (Δ VIX) in our focused model, but results did not significantly varies.

problem, we also follow the variance decomposition proportions method of detecting significant multicollinearity among the explanatory variables (Belsley et al. 1980) and the variance inflation factor (VIF). From the analysis of correlation matrix, variance decomposition proposition method and VIF indicator, we deduce that our proxy factors of default risk do not possess the problem of significant multicollinearity which leads us to evaluate the results of our specification (3), (4) and (5) respectively.²³

Following specification (3), from table (1) column (5), we took only those proxy factors which show statistically significant relation and construct a monthly time series of change in default risk factor delineating the optimal mix effect of statistically significant proxies. In particular, we select²⁴ distance to default as our principal proxy factor (X_{1t} in specification 3) representing optimal default risk indicator.

Figure 1; graphically depict change in default risk post hoc indicator time series (ΔP_{dd}) along with change in yield spread (ΔYS) of US corporate bonds for period of 2000-2010. It shows that our risk of default estimator almost systematically depicts a structural relation with overall yield spread of non-financial US corporate bond securities, showing similar long term trend. Further, our variable of interest (ΔP_{dd}) quite effectively gauge the dot-com (2000-2002) and sub-prime (2007-2009) crises, reporting more than 320 basis points change in the overall risk of default during sub-prime crisis and almost 200 basis points increase in the overall risk of default for US Corporate bond issues in dot-com crisis.

Figure 1: Default risk post-hoc estimators (ΔP_{dd}) and yield spread of US corporate bond securities (ΔYS).

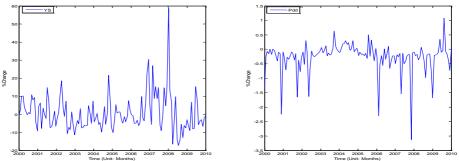


Table (2), column (1) show the results of our specification (4), which is the focused objective of this study i.e. evaluating the effect of optimal risk of default estimator on overall yield spread of US corporate bonds.

Column (1) table (2), shows that our post-hoc default risk estimator is statistically significant in explaining variations in US corporate bond yield spread. In particular, regression coefficient indicates that due to one percentage positive change in default risk estimator (i.e. as it is based on distance to default principal proxy factor, therefore

²³Related results could be provided upon request.

²⁴On the basis of Akaike information criteria (AIC).

it means firm is moving away from default level) the respective yield spread reduces by 0.623 units. In addition, explanatory power of our risk of default indicator is 41 percent which relatively supports theoretical model assumptions in treating risk of default as a primary component in valuing risky corporate securities.²⁵ Furthermore, significant constant term in specification (4) shows presence of non-default factors in overall formulation of corporate bond yield spread. On the other hand B-G test show no evidence of serial correlation in residual error terms. In addition, through Ramsey Reset test, we validate the evidence of delineating specifications based on our surmise of objectively treating default risk as linear combination of significant proxy variables.²⁶ In turn, these results support existing empirical literature of linear relationship between focused dependent and explanatory variables.

Dependent variable is yield spread change	1	2
ΔP_{ddt}	-0.62 ^c	
	(0.33)	
$\Delta P_{ddt \ crisis}$		-1.22 ^a
		(0.46)
$\Delta P_{ddt non-crisis}$		-0.5
		(0.71)
Φ	0.04°	0.64 ^b
	(0.02)	(0.33)
Observations	132	132
R ² Adjusted	0.41	0.48
D-W	1.99	2.05
Breusch Godfrey LM Test		
LM-Statistics	5.90	1.94
P-value	0.93	0.37
Ramsey Reset Test		
F-statistics	0.06	0.76
P-value	0.93	0.46

 Table 2: Default risk post-hoc estimator effect on US corporate bond yield spread

Note: Standard errors are reported in brackets. Column (1) reports effect of *post-hoc* default risk estimator on the change in dependent variable. Column (2) manifests sub-prime financial crisis (2007-2009) effect of our optimal default risk estimator over yield spread. Whereas a, b & c represents significance level at 1%, 5% & 10% respectively.

In addition, table (2) column $(2)^{27}$ report results of specification (5) which delineates the effect of financial system disruptions in the US due to sub-prime crisis (2007-

²⁵Supporting Churm & Panigirtzoglou (2005) and Tang & Yan (2010).

²⁶Here, p-value is more than 5%, so we can not reject null hypothesis of correct linear model specification.

²⁷We deeply thank anonymous reporter comments to include sub-prime financial crisis effect in our specifications.

2009). In particular, results in column (2) outline how significantly our focused optimal indicator of default risk affects yield spread variations in this crisis period. It is really interesting to note that by bifurcating our focused time interval into crisis and non-crisis period, our risk of default estimator shows engaging effect on overall fluctuation in yield spread. It now manifest that statistically significant relation in column (1) is mainly due to the effects of financial crisis of 2007-2009. Table (3) column (2) shows that our default risk estimator is statistically significant during sub-prime crisis period in explaining variations in US corporate bond yield spread.

On the contrary, during normal (i.e. non-crisis) period it is not statistically significant though following presumed relation with our dependent variable. In addition, significance level of our constant term, table (2) column (2), has increased emphasizing the importance of presence of non-default risk factors in overall formulation of US corporate bond yield spread in this crisis period.²⁸ Further, explanatory power of our post-hoc estimator increased by 7 percent (i.e. to 48%), indicating robust effects in explaining abrupt variations in overall yield spread in sub-prime crisis period. In similar vein, results of Breusch Godfrey LM test of serial correlation for residual error term shows no evidence of significant serial correlation. In general, by comparing results in column (1) and (2), table (2), we draw potential outcome that our post-hoc estimator mainly gauge significant relation on overall yield spread in the sub-prime crisis period and its effect dilates across non-crisis period on aggregate basis. This in turn relatively supports theoretical strand of related literature in treating default risk as a primary pillar in the valuation of risky corporate securities.²⁹

5 Conclusion

The aim of this study is to explore the treatment of default risk conundrum between theoretical and empirical proponents of valuing risky corporate debt securities. The dominant view exists in theoretical literature is that risk of default is fundamental to value risky corporate securities which is evident in structural models of valuing credit spread. Whereas major empirical literature on the issue shows that non-default risk factors are more important in valuing these risky corporate securities. Current study attempts to fill this void by presuming that problem to this debate lies in the measurement techniques existing empirical literature use as a subjective treatment of default risk. Using risk of default as an objective linear function of significant proxy factors, we construct an optimal indicator (following Lubotsky and Wittenberg 2006) for default risk which contributes relatively to solve this debate. In particular, our risk of default estimator significantly explains change in US corporate bond yield spread during recent financial crisis (2007-2009). In particular, it explain about 48 percent change in overall yield spread model, indicating dominance of theoretical³⁰ school of thought in treating default risk as a cornerstone in risky corporate debt valuation process.

²⁸Supporting results of Dick-Nielsen, Feldhütter and Lando (2012).

²⁹We also run regressions by introducing dot-com crisis (2000-2002) effect but results were not statistically significant probably showing limitations of our sample that lacks data in this sector. ³⁰Provided by Merton (1974).

References

Altman, E (1968) Discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance. Vol. 23, pages 589-609.

Basel Committee on Banking Supervision (2001) The Internal Ratings Based Approach. http://www.bis.org/publ/bcbsa05.odf.

Belsley, D, Welsch, R and Kuh, E (1980) Regression Diagnostics, Wiley, New York.

Black, F and Cox, J (1976) Valuing corporate securities: some effects of bond indenture provisions. The Journal of Finance. Vol. 31, pages 351-367.

Black, F and Scholes, M (1973) The pricing of options and corporate liabilities. The Journal of Political Economy. Vol. 81, no. 3, pages 637-654.

Chance, D (1990) Default risk and the Duration of Zero Coupon Bonds. The Journal of Finance. Vol. 45, pages 265-274.

Churm, R and Panigirtzoglou, N (2005) Decomposing Credit Spreads. Working Paper no. 253. http://ssrn.com/abstract=724043.

Collin-Dufresne, P, Goldstein, RS and Martin, JS (2001) The Determinants of Credit Spread Changes. The Journal of Finance. Vol. LVI, no. 6, pages 2177-2207.

Couderc, F, Renault, O and Scaillet, O (2007) Business and Financial Indicators: What are the determinants of default probability changes? Working paper. http://www.hec.unige.ch:8080/scaillet/pdfs/BusinessDeterminants.pdf.

Delianedis, G and Geske, R (2001) The Components of Corporate Credit Spreads: Default, Recovery, Tax, Jumps, Liquidity, and Market Factors. Working paper no. 2. http://www.anderson.ucla.edu/x5962.xml.

Dick-Nielsen, J, Feldhütter, P and Lando, D (2012) Corporate bond liquidity before and after the onset of the subprime crisis. Journal of Financial Economics. Vol. 103, pages 471-492.

Duffie, D (1999) Credit swap valuation. Financial Analyst Journal. Vol. 55, no. 1, pages 73-87.

Elton, EJ, Gruber, MJ, Agarwal, D and Mann, C (2001) Explaining the Rate Spread on Corporate Bonds. The Journal of Finance. Vol. LVI, no. 1, pages 247-277.

Ericsson, J and Renault, O (2006) Liquidity and Credit risk. The Journal of Finance. Vol. 61, no. 5, pages 2219-2250.

Fisher, L (1959) Determinants of risk premiums on corporate bonds. The Journal of Political Economy. Vol. 67, no. 3, pages 217-237.

Standard and Poor's, (2011) 2010 Annual US Corporate Default Study and Rating Transitions. Global Fixed Income Research. http://www.standardandpoors.com/ratings/articles/en/us/?articleType=HTML&assetI D=1245302234800. Huang, JZ and Huang, M (2003) How much of the Corporate-Treasury Yield Spread is Due to Credit Risk. Working paper Penn State University. http://forum.johnson.cornell.edu/faculty/huang/Huang-Huang.pdf.

Jones, P, Mason, S and Rosenfeld, E (1984) Contingent Claims Analysis of Corporate Capital Structures: An Empirical Investigation. The Journal of Finance. Vol. 39, no. 3, pages 611-625.

Leland, H and Toft, K (1996) Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. The Journal of Finance. Vol. 51, pages 987-1019.

Liu, S, Shi, J, Wang, J and Wu, C (2009) The determinants of corporate bond yields. The Quarterly Review of Economics and Finance. Vol. 49, pages 85-109.

Longstaff, F and Schwartz, E (1995) A simple approach to value risky fixed and floating rate debt. The Journal of Finance. Vol. 50, pages 789-820.

Longstaff, F, Mithal, S and Neis, E (2005) Corporate yield spreads: Default risk or liquidity? New evidence from the credit default swap market. The Journal of Finance. Vol. 60, no. 5, pages 2213-2253.

Lubotsky, D and Wittenberg, M (2006) Interpretation of regression with multiple proxies. The Review of Economics and Statistics. Vol. 88, no. 3, pages 549-562.

Merton, RC (1974) On the pricing of corporate debt: the risk structure of interest rates. The Journal of Finance. Vol. 29, pages 449-470.

Nakashima, K and Saito, M (2009) Credit Spreads on Corporate Bonds and the macro economy in Japan. Journal of Japanese and International Economies. Vol. 23, no. 3, pages 309-331.

Schuermann, T (2004) What do we know about loss given default, in Shimko, D (2nd ed), Credit Risk: Models and Management, Incisive Media, London, pages 249-274.

Sharpe, W, Alexander, GJ and Bailey, JV (1999) Investments, Prentice Hall College Division, New Jersey.

Tang, DY and Yan, H (2010) Market conditions, default risk and credit spreads. Journal of Banking and Finance. Vol. 34, no. 4, pages 743-753.

Thomson Reuter data base.

Tsuji, C (2005) The Credit spread puzzle. Journal of International Money and Finance. Vol. 24, pages 1073-1089.

Van Landschoot, A (2004) Determinants of euro term structure of credit spreads. Working paper no. 397, European Central Bank, http://www.ecb.int/pub/pdf/scpwps/ecbwp397.pdf.

Weller, CE (2009) Economic Snapshot for August 2009. Center for American Progress. http://www.americanprogress.org/experts/WellerChristian.html/.

Predicting Stock Returns: Evidence from the Istanbul Stock Exchange

Güneş Topçu1

¹ Department of Business Administration, Middle East Technical University, gunes.topcu@hotmail.com

Abstract. This paper examines the predictability of stock returns in the İstanbul Stock Exchange (ISE) by using five types of financial ratios: profitability, leverage, investments, liquidity, and dividend policy. Stock returns were predicted with data of financial statements under the framework of the efficient market hypothesis. The data consist of annual stock returns and financial ratios for the listed firms on the ISE from 1997 to 2011. Generalized method of moments was used in analyzing panel data. The results showed that the accounting data of firms listed in the ISE turned out to possess power in predicting stock returns. This suggests that investors earn abnormal profits by forming investment strategies based on past accounting information in the Turkish stock market.

Keywords: Generalized method of moments (GMM), efficient market hypothesis, portfolio formation

1 Introduction

With the increasing globalization of financial markets, the interests of investors in emerging markets have grown; therefore the predictability of stock returns in emerging markets has become a more important issue than before. Previous literature shows that different prediction methodologies are used in order to predict stock returns in an inefficient market. However, it cannot be concluded with certainty that a particular method outperforms the other. One method of previous research was to use information available from financial statements as a tool for forecasting the stock returns. In this study, by using efficient market hypothesis, the link between the stock returns and financial ratios such as profitability, leverage, liquidity, investment, financial structure, and dividend policy will be investigated for the emerging Turkish stock market. Then, by using this link, this study will form winner and loser portfolios and evaluate their performances.

Market efficiency hypothesis proposed by Eugene Fama in the 1960s is used in this research in finding a relationship between stock returns and financial ratios. If the markets are efficient, it is impossible to gain excess returns by using past accounting information. When reviewing the past research, I find many studies on the issue of market efficiency in Turkish stock market such as Buguk and Brorsen (2003), Müslümov et al. (2006), Yalama and Çelik (2008), and Korkmaz and Akman (2010). Korkmaz and Akman (2010) tested the weak-form efficiency of the ISE market by using unit root and co-integration tests, and concluded that the ISE is weakly efficient. In contrast, Yalama and Celik (2008) claimed that money market has semi-strong form of efficiency, but capital market does not by using Toda Yamamoto Causality test. Unlike previous studies, several testing methodologies were used in the study by Buguk and Brorsen (2003): (1) Augmented Dickey-Fuller test, (2) Geweke and Porter- Hudak (GPH) fractional integration, (3) Lo and MacKinlay (LOMAC) variance ratio test, and (4) a modified variance ratio test. Three out of the four test results showed that the random walk hypothesis could not be rejected, which is interpreted as evidence for market efficiency. In addition, the study by Müslümov et al. (2006) stands out in that the study tested not only whether the Turkish stock market is weakly efficient, but also whether the market is becoming more efficient. By dividing the data of the weekly return series of ISE-100 index from 1991 to 2001 into two equal parts, they tested market efficiency. They found that the first part of the data did not support random walk behavior, but the second part which contains more recent data showed evidence for weak-form efficiency. The differences often found in the results of the previous studies can be probably attributed to the use of different statistical methodologies used in the analysis of the data.

The motivation of this research is as follows. Compared to the abundance of research about developed markets, not much research has existed about the stock market efficiency in Turkey, despite the above mentioned research. In particular, not many studies which used accounting information are found for the emerging markets such as Turkey1. This study attempts to fill this gap by examining the predictability of stock returns in the emerging Turkish stock market by using financial ratios. Considering the previous results, this study will conduct econometric tests on the assumption that the Turkish Capital Market is weakly efficient. To the best of author's knowledge, seldom research could be found that included dividend ratios such as dividend payout ratio and dividend yield in this type of research on the ISE.

The paper is organized as follows. Section 2 summarizes the previous studies. Section 3 presents the data and the research method. Section 4 shows summary statistics and the empirical evidence. The summary and the conclusion will be provided in Section 5.

¹See Aktaş (2008), Horosan (2009), and Büyükşalvarcı (2011).

2 Literature Review

Ball and Brown (1968) investigated whether security prices reflect the flow of information in the US market as it becomes available by using event study methodology. Data consisted of three classes, namely, the contents of income reports over the period 1946 to 1966, the dates of the report announcements over the period 1957 to 1965, and the movement of security prices around the announcement dates over the period 1946 to 1966. They concluded that accounting earnings and some of its components capture information *id est* contained in stock prices. Accounting earnings were valued positively by investors.

Ou and Penman (1989) used accounting ratios that describe a firm's leverage, activity and profitability to perform a financial statements analysis. The future earnings implications of current accounting data were calculated by using a logit model that summarizes the large number of accounting ratios into a single value which indicates the direction of one-year-ahead earnings changes. Sample covered the period 1965 to 1972. They concluded that financial statements data predicts successfully future earnings performance.

Elleuch (2009) examined whether stock returns can be predicted by using fundamental analysis based on historical accounting information. Sample consisted of the firms traded on the Tunusian stock market over the period 1995 to 2001. He chose 12 indicators for the analysis: inventory, accounts receivable, investments, gross margin, labor force, return-on-assets, cash flow, accruals, leverage, liquidity, and assets turnover. Elleuch calculated the Spearman correlations between firm's profitability, the twelve indicators, and the aggregate fundamental score, i.e. F_score which shows the sum of the individual binary variables specified for the 12 accounting signals. According to the results, there was a significant positive correlation between the F_score and returns. He concluded that historical accounting signals could be used to predict stock returns.

Penman, Richardson and Tuna (2007) investigated stock returns and financial statement data relationship for the US stock market from 1962 to 2001. Financial statement data included levered and unlevered book-to-price ratios and market leverage. They used a regression model to estimate the significance of the variables. They found that the enterprise book-to-price ratio is positively related to returns and conditional upon the enterprise book-to-price ratio, market returns are negatively correlated with subsequent returns. Moreover, they found a negative relation between expected returns and leverage.

Mcmanus, Gwilym and Thomas (2002) examined the relationship between stock returns and dividends in the context of the UK stock market over the period 1958 to 1997. They employed generalized method of moments. They concluded that behind dividend yields, payout ratio is an important factor in explaining stock returns.

Pontiff and Schall (1997) used book-to-market ratio of the Dow Jones Industrial Average (DJIA) to predict stock returns over the period 1926 to 1994. They estimated

the first-order autoregressive process for book-to-market ratio, interest yields, and dividend yields. They found that DJIA book-to-market ratio contained information about future returns that was not contained in other variables such as interest yields and dividend yields. The possible reason for the explanation power of book-to-market ratios seemed to result from the relationship between book value and future earnings.

Lamont (1998) examined the relationship between earnings and expected returns of stocks by using postwar US data over the period 1947 to 1994. Data consisted of quarterly excess returns, dividends, and earnings. He used ordinary least squares (OLS) to estimate equations. According to the estimation results, the coefficient for the dividend yield was significantly positive and the coefficient on the earnings yield was significantly negative in a bivariate regression which means that dividend yield and earnings yield seem to predict short-term returns. However, these variables could not explain the expected returns for long horizons.

Bali, Demirtaş and Tehranian (2008) analyzed the predictability of stock returns by using market, industry, and firm-level earnings of the US market. Data were retrieved from the Center for Research in Securities Prices (CRSP) file. They formed all variables in a quarterly basis. The overall data consisted of 289,958 quarterly observations. They used OLS to predict excess market returns and corrected the biases in the parameters by employing the randomization technique of Nelson and Kim (1993). Contrary to Lamont (1998), they found that at the aggregate level, the information content of firm-level earnings about future cash flows was diversified away and neither dividend payout ratio, nor the level of aggregate earnings could forecast the excess market return.

Muradoğlu and Whittington (2001) used the ability of leverage ratios to predict stock returns in the long run for UK stock market. They used "gearing ratio", i.e. the ratio of total debt to total financing of the firm to represent the leverage of the companies in the sample data. The study found that over the period 1990 to 1999, debt ratios could have been used to predict share price out-performance in the UK.

Alexakis, Patra and Poshakwale (2010) examined the predictability of stock returns for the Greek Stock Market by using 10 financial ratios over the period 1993 to 2006. They employed a generalized method of moments framework to estimate the dynamic panel data regression model. They found that the selected financial ratios contain significant information to predict stock returns, and Greek market was not semi-strong form efficient.

3 Data and Methodology

3.1 Data Description

The data used in this study contain the annual stock prices, dividends, and financial ratios for a sample of 36 representative Turkish firms listed on the ISE. In this study, the ISE 100 index is used as the basic index for the ISE Stock Market. The ISE 100 Index, one of the major indices of the ISE, comprises the stocks of 100 selected companies listed on the National Market and the Corporate Products Market as well. Monthly returns of the Istanbul Stock Exchange 100 index over the period January 1997 to September 2011 is given in figure 1.

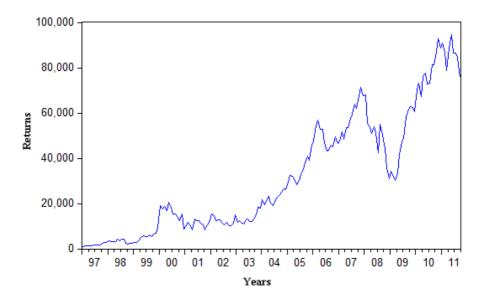


Figure 1: Monthly returns of the Istanbul Stock Exchange (ISE) 100 index

This paper used data of non-financial firms because (1) it is difficult to use the data of financial firms together with those of manufacturing firms and (2) the leverage levels of financial firms are known to be higher compared to those in other manufacturing industries. Data for this study included the prices, dividends, and the financial ratios from January 1997 to September 2011 partly because of availability of the adjusted stock prices. Considering the previous studies on the evaluation of investment performance, this study has selected five types of financial ratios such as profitability, leverage, liquidity, investment, and dividend policy. Table 1 provides a brief summary of the selected financial ratios.

Variables such as return on equity (ROE) and net profit margin (NPM) are used as proxy variables for profitability. A positive relationship is expected between

profitability and stock returns because the more profitable a firm becomes, the higher its stock price becomes. The literature such as the proposition two of Modigliani-Miller theorem shows a positive relationship between the leverage and the stock returns as the states.² Another claim for the positive sign can be stated like this: The higher the debt to equity ratio, the higher the return on equity. As a liquidity ratio, current ratio is used in this study to measure a company's short-term solvency. A high current ratio shows that the firm is sound enough to pay its short-term obligations, and it will result in the increase in stock returns. The variables such as price-to-earnings ratio (P/E) and price-to-book value (PBV) are used as proxy variables as investment indicators. Investors use P/E ratio to determine what the investors get when they purchase stocks. The previous studies have found a negative relationship between the stock returns and the P/E ratio.³ The variable price-to-book value (PBV) shows whether a company is over or undervalued in the market. If a company has a low PBV, it is considered a good investment opportunity. Aydogan and Gursoy (2000) asserted that both price-to-earnings ratio and price-to-book value appear to predict stock returns; however, explanatory powers of these two ratios turned out to be very low in shorter return horizons.

 Table 1: Financial Ratios

Ratio	Formulae
Profitability ratios	
Return on assets (ROE) Net profit margin (NPM)	(Net income / shareholders' equity) *100 (Net income / net sales)*100
Leverage ratios	
Debt to Equity (DE)	(Total debt / total equity)*100
Liquidity ratios	
Current ratio (CR)	(Current assets / current liabilities)*100
Investment ratios	
P/E ratio (PER)	Stock price / earnings per share
Price to book value (PBV)	Stock price / book value per share
Dividend policy ratios	
Dividend Payout Ratio (DPR)	(Dividends / net income)*100
Dividend Yield (DY)	(Dividends per share / stock price)*100

² see Korteweg (2004), Muradoğlu and Whittington (2001), and Miller and Modiglianni (1958).

see Campbell and Schiller (1998) and Trevino and Robertson (2002).

As a proxy for dividend policy ratios, dividend yield shows how much money is paid back as dividends at the end of the period from shares purchased at the beginning of the period. It is different from dividend payout ratio in that the variable dividend yield shows a direct relationship between the stock price and the dividend.

3.2 Methodology

In this study, I used panel data analysis because of its several advantages: (1) The information from time series is compounded with the information from cross-sectional units; (2) The addition of cross sectional variation to time series variation increases estimation efficiency: Since standard errors decrease, we get higher t-ratios (Erlat and Ozdemir, 2003), and (3) Degrees of freedom will be increased with the construction of panel data which will reduce the collinearity among explanatory variables (Hsiao, 2003).

I used the following dynamic panel data regression model that is used in Alexakis et al.'s study:

$$\mathbf{R}_{it} = \delta \mathbf{R}_{i, t-1} + \mathbf{X}_{it}^{*} \boldsymbol{\beta} + \mathbf{u}_{it} \quad i = 1, ..., 36; \qquad t = 1, 14$$
(1)

where

$$u_{it} = \alpha_i + \varepsilon_{it} \tag{2}$$

In equations (1) and (2), firms are indicated by *i* and time by *t*. X_{it} is the vector of explanatory variables and u_{it} is the error term where α_i and ε_{it} are independently and identically distributed and uncorrelated with each other. Different from Alexakis et al.'s study, I included the dividends variable to calculate returns.

The realized log gross return for each firm in year t is denoted as R_{ii} which can be written as

$$ln (1+R_{it}) \equiv \ln(P_{it}+D_{it}) - \ln(P_{it-1})$$
(3)

In equation (3), P_t is the real price of a stock measured at the end of time period t and D_t is the real dividend paid on the stock during period t.

Since the model includes a lagged dependent variable which creates a correlation between itself and the error term, the fixed effects estimator becomes inconsistent for a small T. This bias does not vanish even though the number of the individuals increases (Nickel 1981). As Baltagi (2008, p. 148) states, the random effects estimator is also biased in a dynamic panel data model because of the assumed correlation between the quasi-differenced values of the lagged dependent variable and the quasi-differenced error term.

A dynamic version of the panel data analysis, generalized method of moments (GMM) proposed by Arellano and Bond (1991), was used in order to estimate the regression equation because of the potential endogeneity in the model and the presence of the lagged dependent variable. The GMM estimator has been developed for large N and small T panels. Therefore, it is appropriate to use GMM method for this dataset where N=36 and T=14. The method of moments basically includes the estimation of the parameters of the model in question by imposing the restrictions assumed about the moments of the variables involved, to the sample moments, and it becomes generalized when the number of moment conditions is greater than the number of parameters being estimated (Hansen 1982). The lagged values of the dependent and predetermined independent variables are used as instrumental variables in the first differenced model at a given period in the Arellano and Bond's dynamic panel GMM method. Elsas and Florysiak (2010) assert that the first differenced GMM estimators are biased since they ignore the fact that the debt ratios are bounded, i.e. they are between 0 and 1. Although first differencing the equation as it is done in this study eliminates the individual effects, it may create a flaw for the debt ratio and may lead to biased results.

Using instrumental variables makes independent variables strictly endogenous. I checked the validity of the instrumental variables by using the Sargan's test of overidentifying restrictions (1958, 1988). I used balanced panel because the dataset for any given period has no missing values.

I first conducted unit root test in order to confirm stationarity of the variables before beginning the estimation process. In the literature, several unit root tests are found for the panel data such as Levin, Lin and Chu (LLC) (2002), Im, Pesaran and Shin (IPS) (2003), Choi (2001), and Maddala and Wu (1999). In this study, I used LLC and IPS tests where null hypotheses are the existence of unit roots. Although they both evaluate the null hypothesis of unit roots, their null and alternative hypotheses are different in that in IPS approach, the rejection of the null hypothesis implies that all series is stationary.

4 Empirical Results

4.1 Unit Root Test Results

The results of the two panel unit root tests are given in Table 2. I note that both the LLC and IPS tests reject the null hypothesis of unit root. We find in this table that the null hypothesis of unit root is rejected for the majority of the variables. The LLC test results seem to be largely consistent with the IPS test results. The results showed that all independent variables are stationary except variables such as the current ratio and the price to book ratio. Those two variables also became stationary when they are first differenced.

-	L	evel IPS and LLC	First Di Test Results	fference
Variabl es	IPS		IPS	LLC
CR	-0.697	0.407	-13.590***	- 12.412***
DE	-3.821***	-3.6153***		
DPR	-8.887***	-11.479***		
DY	-3.263***	18.137***		
NPM	-7.762***	-8.203***		
PBV	-0.202	-1.460***	-5.677***	
PER	-7.761***	-5.706***		
ROE	-9.135***	-8.665***		
RT	-21.156***	-19.222***		

 Table 2 : Panel Unit Root Test Results

2. "***" : significant at the 1% level.

4.2 GMM Estimation Results

Table 3 reports estimation results. We find in this table coefficients, standard errors, and t-statistics for the financial ratios for two different periods: 1997 to 2009 and 1997 to 2010. Consistent with the literature, results indicate a significant and positive relationship between the returns and the three ratios: the dividend yield, the return on equity, and the current ratio. The dividend yield turned out to have the highest explanatory power of the stock returns among the three ratios. The positive relationship between the dividend yield and the returns shows that dividend yields can be used as a decision criterion for the stock investments.

	1997-2009		1997-2010		
Variables	Coefficient	t-statistics	Coefficient	t-statistics	
CR	0.074*	2.131	0.064*	2.011	
DE	0	0.092	-0.001	-0.59	
DPR	0.068	0.359	0.143	0.712	

Table 3: GMM Estimation Results

DY	0.382* 3.229		0.375*	3.082	
NPM	0.192 1.047		0.158	1.013	
PBV	0.001 0.647		0.001	0.806	
PER	0 0.203		0	0.039	
ROE	2.084* 2.661		1.802*	2.336	
J-statistic	16.93		15.822		
Sargan Test (p-value)	0.075		0.105		
SE of regression	0.566		0.565		

Note: "*" : significant at the 5% level.

The relationship between the stock return and the ROE was estimated to be positive as the theory predicts. The logic can be understood in this way: as the firm becomes more profitable, its share price will increase as well. In addition, current ratio turned out to be also positive at the 5% significance level as the theory predicts. Other financial ratios turned out to be statistically insignificant.

By using the Sargan test, I also tested the validity of instrument sets used in this panel estimation. The null hypothesis in the Sargan test is this: the excluded instruments are uncorrelated with the error term. In other words, it is a test of overidentifying restrictions. Considering p-values such as 0.105 and 0.075, I failed to reject the null hypothesis. This indicates that the instrument sets are valid.

4.3 An Application to Investment Strategy

As an application step, I applied the previous GMM results to the issue of investment management. I followed the investment strategy that is used in Alexakis et al.'s (2010) study. I predicted returns for the 2010 by using the estimation results of 1997 to 2009. Next, the stocks are ranked according to their estimated returns. A winner portfolio was formed from the 10 stocks with the highest returns while a loser portfolio was formed from the 10 stocks with the lowest returns. I repeated the calculations for the portfolios of the year 2011 by using the period 1997 to 2010 for the estimated coefficients. However, I had to use limited data of the nine month for calculation because, at the time of writing this study, the annual financial statements for 2011 have not yet been revealed. In order to maintain integrity of the data, I used variables available for one year from September 2010 to September 2011. The results are given in Table 4. I find in this table that winner portfolios' performance outperformed that of the loser portfolio in 2010 and 2011. This implies that investors may get abnormal profits from the following trade by selling the stocks in the loser portfolio.

Actual Annual Returns	2010	2011
Winner Portfolio	4.72	0.327
Loser Portfolio	1.95	-2.288

5 Conclusion

I examined the relationship between the stock returns and the financial ratios over the period 1997 to 2011. Generalized method of moments was used in order to estimate the equation for the balanced panel dataset. The results showed that with financial ratios, it was possible to explain stock returns in the Istanbul Stock Exchange. As an application to investment strategy, I also formed winner and loser portfolios based on the estimation results. The results suggest that in developing countries such as Turkey, financial markets are not efficient in a semi-strong form sense, implying that investors have opportunities of gaining excess returns by using past accounting information.

References

Aktaş M. (2008) Istanbul Menkul Kıymetler Borsasında Hisse Senedi Getirileri ile İlişkili Olan Finansal Oranların Araştırılması. *Istanbul University Journal of the School of Business Administration*. 37,2, 137-150.

Alexakis, C., P., Theophano, and P., Sunil. (2010) Predictability of Stock Returns Using Financial Statement Information: Evidence on Semi-strong Efficiency of Emerging Greek Stock Market. *Applied Financial Economics*. 20, 1321-1326.

Arellano, M. and S. Bond (1991) Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*. 58, 277–97.

Aydoğan K. and G. Gürsoy (2000) P/E and Price-to-Book Ratios as Predictors of Stock Returns in Emerging Equity Markets. *Emerging Markets Quarterly*. 4, 60-67.

Bali, G., Turan, K. Özgür Demirtaş and Hassan Tehranian (2008) Aggregate Earnings, Firm-Level Earnings and Expected Stock Returns. *Journal of Financial and Quantitative Analysis*. 43,3, 657-684.

Ball, R. and P. Brown (1968) An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research*. 6, 159-178.

Baltagi, B. H. (2008) *Econometric Analysis of Panel Data (Fourth Edition)*, Wiley, New York, NY.

Büyükşalvarcı A. (2011) Finansal Analizde Kulalanılan Oranlar ve Hisse Senedi Getirileri Arasındaki İlişki: Ekonomik Kriz Dönemleri İçin İMKB İmalat Sanayi Şirketleri Üzerine Ampirik bir Uygulama. *Atatürk University Journal of Economics and Administrative Sciences*. 25,1, 225-240.

C. Buguk and B. W. Brorsen (2003) Testing Weak-form Market Efficiency: Evidence from Istanbul Stock Exchange. *International Review of Financial Analysis*. 12, 5, 579–590.

Campbell, J. and R. Shiller (1998) Valuation Ratios and the Long-Run Stock Market Outlook. *Journal of Portfolio Management*. 24, 11-26.

Cheung, K. J., R. Chung and J.B. Kim (1997) The Profitability of Trading Strategies Based on Book Value and Earnings in Hong Kong: Market Inefficiency Versus Risk Premia. *Journal of International Financial Management and Accounting*. 8, 204–33.

Choi, I. (2001) Unit root tests for panel data. *Journal of International Money and Finance*. 20, 249–72.

Elleuch, J. (2009) Fundamental Analysis Strategy and the Prediction of Stock Returns. *International Research Journal of Finance and Economics*. 30, 95-107.

Elsas, R., and D. Florysiak (2010) Dynamic Capital Structure Adjustment and the Impact of Fractional Dependent Variables. Working Paper, University of Munich.

Erlat, H., N. Özdemir (2003) A Panel Approach to Investigating the Persistence in Turkish Real Exchange Rates. *Topics in Middle Eastern and North African Economies.* E. M. Cinar, 5.

Hansen, L.P. (1982) Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*. 50, 1029-1054

Horasan, M. (2009) Fiyat/Kazanç Oranının Hisse Senedi Getirilerine Etkisi: İMKB 30 Endeksi Üzerine Bir Uygulama. *Journal of Economics and Administrative Sciences*. 23,1, 182-192.

Hsiao, C. (2003) Analysis of Panel Data (Second Edition), Cambridge University Press, Cambridge.

Im, K.S., M.H. Pesaran and Y. Shin (2003) Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics*. 115, 53-74.

Korkmaz, M. and G. Akman (2010) Testing the Weak Form Efficiency on İstanbul Stock Exchange. *Trakia Journal of Sciences*. 8, 3, 39-50.

Korteweg, A. (2004) Financial Leverage and Expected Stock Returns: Evidence from Pure Exchange Offers. Working Paper, Stanford Graduate School of Business.

Lakonishok, J., A. Shleifer and W. R. Vishny (1994) Contrarian Investment, Extrapolation, and Risk. *Journal of Finance*. 49, 1541–1578.

Lamont, O (1998) Earnings and Expected Returns. *Journal of Finance*. 53, 1563-1587.

Levin, A., C.F. Lin and C.S.J. Chu (2002) Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties. *Journal of Econometrics*. 108, 1-24.

Maddala, G. S. and S. Wu (1999) A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics*. 61, 631–52.

Martinez, I. (1999) Fundamental and Macroeconomic Information for the Security Prices Valuation: the French Case. *Managerial Finance*. 12, 17–30.

Modigliani, F. and M. Miller (1958) The Cost of Capital, Corporation Finance and the Theory of Investment. *American Economic Review*. 48, 261-297.

Muradoglu, G. and M. Whittington (2001) Predictability of Stock Returns by Using Debt Ratios. Working Paper, City University Business School, London, United Kingdom.

Nelson, C.R., AND M.J. Kim (1993) Predictable Stock Returns: The Role of Small Sample Bias. *Journal of Finance*. 48, 641-661.

Nickel, S.J. (1981) Biases in dynamic models with fixed effects. *Econometrica*. 49, 1417–1426.

Ou, A. J. and H. S. Penman (2002) Financial Statement Analysis and the Prediction of Stock Returns. *Journal of Accounting and Economics*. 11, 4, 295-329.

Penman, H.S., S.A. Richardson and I. Tuna, (2007) The Book-to-Price Effect in Stock Returns: Accounting for Leverage. *Journal of Accounting Research*. 45, 2, 427-467.

Pontiff, J. and L.D. Schall (1998) Book-to-Market Ratios as Predictors of Market Returns. *Journal of Financial Economics*. 49, 2, 141-160.

Sargan, J. D. (1958) The Estimation of Economic Relationships Using Instrumental Variables. *Econometrica*, 26, 393-415.

Sargan, J.D. (1988) Contributions to Econometrics, Cambridge University Press, Cambridge.

Thomas, S.H., I.D. Mcmanus, and A. O. Gwilym (2002) A Robust Estimation of the Relation between Stock Returns, Size, Dividend Yield and Payout Ratio. *Journal of Business Finance & Accounting.* 31, 9-10, 1355-1387.

Trevino R. and F. Robertson (2002) P/E Ratios and Stock Market Returns. *Journal of Financial Planning*. 15, 76–84.

Yalama, A. and S. Çelik (2008) Financial Market Efficiency in Turkey: Empirical Evidence from Toda Yamamoto Causality Test. *European Journal of Economics, Finance and Administrative Sciences.* 13, 88-93.

Crisis on the Hungarian Government Bond Markets: is that a Liquidity Problem?

Zoltán Monostori1

¹ Department of Finance, Corvinus University of Budapest, monostori.z@gmail.com

Abstract. In the paper, I examine the Hungarian government bond market's liquidity developments in the recent years. Firstly, I explain the importance of market liquidity for central bankers. I identify the most significant economic shocks and their impacts on the market by using various market indicators. The changes in the Hungarian pension system strongly affected the ownership structure of the government bond market, and raised the amount held by non-residents. A simple yield decomposition shows that while over the crisis of 2008-09, the Hungarian sovereign bond yields were enhanced principally by the increase of the credit and liquidity risk premia, the present crisis might strongly increase credit risk premium, but increase liquidity risk premium less significantly.

Keywords: Bond, Liquidity, Index, Ownership, Decomposition

1 Introduction

More and more countries have to face the problem that their government debt / gross domestic product quotients dynamics is not sustainable. The most important factors in this process are smaller growth, a bad structural balance of the budget, and the high financing costs of the government debt, which is related to the increasing sovereign yields. These three factors are closely related, but we can mention that, on the one hand, raising the growth rate and balancing the budget could be done by using either different (e.g., positive fiscal stimulus vs. fiscal tightening) or very unpopular measures (such as making more flexible working laws, or raising the retirement age). On the other hand, it is possible to reach relative quick success in decreasing the sovereign yields by using conservative, predictable fiscal and monetary policy, and offering a good and calm inventorial atmosphere for financial portfolio-capital investors. The focus of the paper is on the kind of risks which affect the fixed income yields, and on the type of risk which is the most costly for the bond issuer, the state. Is

there a liquidity problem on the Hungarian sovereign bond markets, or is this crisis more about solvency doubts? With which indices can we follow the market liquidity, and what were the main shocks in liquidity in the past few years?

The structure of my paper is as follows. First, I start with an introduction about the importance of market liquidity for central bankers. Next, I show different market indicators for the Hungarian sovereign bond markets liquidity, with which I identify the most significant shocks of the past few years, and their impacts on the market. I present one of the shocks, the pension reform more detailed, because of its strong effect on the ownership structure of the bond market. Finally, I show a simple decomposition of the bond yields.

2 The Importance of Liquidity for Central Bankers

Why is market liquidity important for central bankers? Páles and Varga summarized this in three points. First, a liquid money market is necessary for efficient monetary operations. Second, the Monetary Council relies on the information content of the prices of different financial assets when making its monetary policy decisions, which is more difficult in the case of low liquidity when the exchange rate, interest rate and inflation expectation estimations derived from traded assets are inexact. Third, in the case of relatively weak liquidity, the effects of different financial and economic shocks can be stronger (Páles and Varga 2008).

The literature uses the term liquidity in many different meanings. Nikolaou defines three types of liquidity concerning the financial systems. First, the central bank liquidity, which is the ability of the central bank to provide liquidity for the financial system if it is necessary. Second, financing liquidity, which is the ability of the commercial banks to open the positions they want when their liabilities run out. Third, market liquidity (Nikolaou 2009). While examining market liquidity, we measure whether large volume transactions can be carried out in the given financial market within a short period of time and without a significant change in market prices (Bank of International Settlements, 1999). The risk premia in the fixed income market could be increased by the lack of market liquidity, therefore the indicators discussed in the paper quantify the trends in market liquidity.

3 Market Indicators for the Hungarian Sovereign Bond Markets

The difficulties concerning the security-borrowings and the naked short sellings, together with the absence of the repo market are important fields explaining the lack of liquidity on the Hungarian sovereign bond markets.

Market liquidity includes several characteristics of the order book¹. Csávás and Erhart estimate market liquidity in five different dimensions. First, tightness is a dimension about the transaction costs. The typical purpose of measuring tightness is

¹ Order book is the list of orders that have not yet been executed.

the difference between the best available bid and ask prices. The measurement of the bid/ask spread is challenging on the Hungarian fixed income market, since most of the transactions are made on the OTC (over-the-counter) markets. The second dimension, depth, is about an amount: what is the highest transaction amount that can be made without changing the market price? A market is deep if there are many bidand ask prices close to the market price in the order book. The third dimension, breadth, is closely interconnected with depth, and can be seen as a more general version of depth. Many authors do not differentiate between these two dimensions. Breadth also shows something about the number of bid and asks for offers: it measures how many different offers are active in the order book to a specific price. Therefore, in the case of breadth, not only the offers near the market price matter, but also the offers on more extreme prices. The fourth dimension, resiliency is about time: how fast the (probably new) market price is stabilized after a big transaction. The fifth and last dimension is immediacy, which is also a time dimension: how long it takes to make a transaction. The amount of immediacy depends on the number and heterogeneity of the market participants, but also on how developed the trading system is (Csávás and Erhart 2005).²

² Another dimension of market liquidity sometimes discussed in the literature is the (daily) turnover, see (Butler et al. 2005)

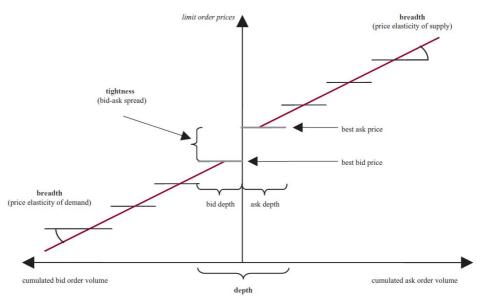


Figure 1: Market liquidity dimensions in a stylised order book

Source: (Csávás and Erhart 2005)

After the publications of the liquidity index of the Bank of England (Bank of England, 2007) and the European Central Bank (European Central Bank, 2007), the MNB (Magyar Nemzeti Bank, the Central Bank of Hungary) has also created a liquidity index that is based on several financial markets sub-indices (EUR/HUF spot FX market, USD/HUF FX-swap market, interbank unsecured money market, and sovereign bond market). The sovereign bond market sub-index has four "sub-sub-indices": bid-ask spread index to measure tightness, return-to-volume index to measure a mixture of resiliency and market depth, transaction size index and number of transactions index to measure market depth. Each of the four indicators can be seen in Figure 2 (Páles and Varga, 2008)³.

³ All of these were calculated such that a raise in the value of an indicator shows a raise in liquidity (for example, the value of the bid/ask spread indicator raises when the bid/ask spread drops, which means the market becomes tighter). To sum the calculated time series, they have to be on a common denominator. Because of the different units of measurement and magnitudes, the series are normalized, so the units of the data become comparable and lose their original unit of measurement.

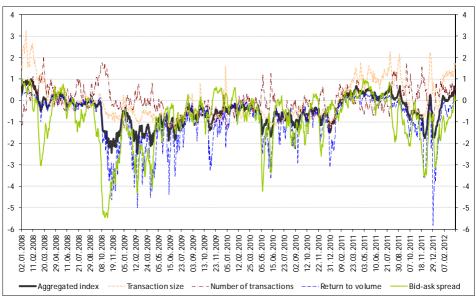


Figure 2: Liquidity indicators of the Hungarian sovereign bond market

As Figure 2 shows, several liquidity shocks have happened during the past years: turbulences on the government bond market in the fall of 2008, drying up of the government bond market in the 3Q 2008, recession in CEE-countries in the 2Q 2009, escalation of the Greek crisis in April and May, 2010, shocks caused by announcements of politicians in June, 2010, announcement of the pension-system's reform in November, 2010, bigger problems in the Euro zone in summer and autumn, 2011, and a country-specific crisis in January, 2012.

4 Changes in the Ownership Structure of the Hungarian Government Bond Market

Hungary also had a big change in the structure of the sovereign bond market that could be linked to the pension reform, which caused a fall in outstanding volume of government bonds, but also abolished (almost absolutely) one of the biggest and most stable sectors on the demand side of the bond market.

Source: MNB

Table 1: Debt securities issued by central government with breakdown by holding sectors (HUF-denominated securities, original maturity over 1 year. Quarterly closing positions at market value, HUF billion)

2011. Q1	2011. Q2	2011. Q3	2011. Q4
57,5	47,6	51,9	40,5
278,7	235,6	238,8	180,8
2 643,0	2 522,1	2 614,5	2 430,5
260,6	242,2	225,5	218,9
118,8	105,1	101,6	106,5
2 657,1	1 428,8	1 409,3	1 342,3
5 958,2	4 533,8	4 589,7	4 279,0
4,7	4,7	3,8	4,5
7,6	5,8	6,7	4,0
0,0	0,0	0,0	0,0
12,3	10,4	10,5	8,5
362,1	343,9	353,5	349,1
6,4	5,2	4,9	7,2
2 588,7	2 851,8	3 242,6	3 014,5
8 985,2	7 792,7	8 253,1	7 698,9
	57,5 278,7 2 643,0 260,6 118,8 2 657,1 5 958,2 4,7 7,6 0,0 12,3 362,1 6,4 2 588,7	57,5 47,6 278,7 235,6 2 643,0 2 522,1 260,6 242,2 118,8 105,1 2 657,1 1 428,8 5 958,2 4 533,8 4,7 4,7 7,6 5,8 0,0 0,0 12,3 10,4 362,1 343,9 6,4 5,2 2 588,7 2 851,8	57,5 47,6 51,9 278,7 235,6 238,8 2 643,0 2 522,1 2 614,5 260,6 242,2 225,5 118,8 105,1 101,6 2 657,1 1 428,8 1 409,3 5 958,2 4 533,8 4 589,7 4,7 4,7 3,8 7,6 5,8 6,7 0,0 0,0 0,0 12,3 10,4 10,5 362,1 343,9 353,5 6,4 5,2 4,9 2 588,7 2 851,8 3 242,6

As Table 1 shows, before the pension reform, the ownership structure of the government bond market was largely dominated by three sectors: credit institutions, insurance companies together with pension funds, and the rest of the world.

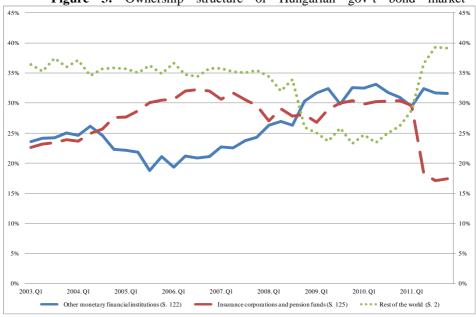


Figure 3: Ownership structure of Hungarian gov't bond market

Source: MNB

As it can be seen in Figure 3, the change in the past years in the mentioned three sectors was significant, and in periods of high risk-aversion, non-residents sold their bonds, which had a pro-cyclical effect on the business cycle. Therefore, the present rising market share of non-residents could pose a risk in the future.

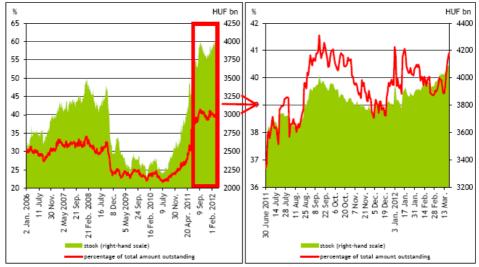


Figure 4: Government securities held by non-residents and the proportion to the total amount outstanding

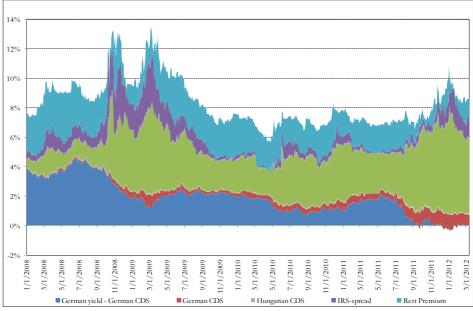
Source: MNB

5 A Simple Decomposition of the Hungarian Government Bond Yields

In a small, open, net borrower economy like in Hungary, the yield curve is determined by the risk free return, the expected change in exchange rates, and the premium expected by the investors in return for the risk they run. The premium is affected by four major types of risks: credit risk, liquidity risk, currency risk and interest rates risk. I show a simple estimation of the premia paid for these risks. While I find reasonable proxies (CDS spreads and interest rate swap spreads⁴) to estimate the credit risk and liquidity risk premia, I will consider only the combined effect of the other two risk types and the exchange rate expectations. I define a combined yield called "rest premium", which incorporates the expected change in exchange rates and the risk premia of the exchange rates and interest rates. It can be estimated as a residual (Monostori, 2011). Finally, I use German yield minus German CDS-spread to represent the risk free rate, and German CDS-spread to represent something I call the "European apocalypse premium".

⁴ Measuring liquidity risk is rather straightforward, unlike the measurement of the liquidity premium, because risk indices are difficult to convert to a premium. What I call IRS-spread is bond yield minus IRS-yield (interest rate swap yield) on the same time horizon. The usage of this is based on the fact that interest rate swaps do not contain liquidity premium, which can easily be accepted since liquidity is not necessary to bid an IRS, and since it is easy to close an open interest rate position by biding another IRS in the different direction.

The estimation of the liquidity premium could be the most questionable. The literature makes a difference between on-the-run and off-the-run liquidity premia. Fleming showed empirical evidence that the not-anymore-auctioned bonds have lower prices compared to others (Fleming, 2003). This does not make a big difference on the Hungarian sovereign bond markets because of the frequent re-auctions of the bonds maturing on the same date. It is also worth mentioning that liquidity premium does not only include information about the present market liquidity, but also an expectation about the market conditions in the future (and there might also be an influence of the current price of liquidity, which is changing in time, as written by Turner (Turner, 2009)).





Sources: Bloomberg, Datastream, Reuters

The above shocks can also be easily found in Figure 5, using the decomposition.

6 Conclusion

The original question is if there is a liquidity problem on the Hungarian sovereign bond market. Liquidity can be measured in many ways, but most indices show that the liquidity of this market (after some drying out in January, 2012) has returned to the long term average, although one of the biggest and most stable sectors was almost absolutely abolished. The decomposition shows that the increase of the yield is mostly caused by the enduring raise of credit risk premium. To sum up we can say that the present crisis might strongly increase credit risk premium, but increase liquidity risk premium less significantly.

7 References

Bank of England (2007) Financial Stability Report. April 2007. p. 18.

http://www.bankofengland.co.uk/publications/fsr/2007/fsrfull0704.pdf.

Bank of International Settlements (1999) Market Liquidity: Research Findings and Selected Policy Implementations. *Committee on the Global Financial System Publications* No. 11th May, 1999.

Butler, A., Grullon, G. and Weston, J. (2005) Stock market liquidity and the cost of issuing equity. *Journal of Financial and Quantitative Analysis* 40: 331–348.

Csávás, Cs. and Erhart, Sz. (2005) Are Hungarian financial markets liquid enough?. *MNB Occasional Papers*. 44.

European Central Bank (2007) *Financial Stability Review*. June 2007: pp. 81-83. http://www.ecb.int/pub/pdf/other/financialstabilityreview200706en.pdf.

Monostori, Z.(2011) *Decomposition of the five year Hungarian fixed income forint yields*. TDK.

http://publikaciok.lib.uni-corvinus.hu/publikus/tdk/MZ_20100330031526.pdf

Nikolaou, K. (2009) Liquidity (risk) concepts definitions and interactions. *European Central Bank Working Paper* No. 1008.

Páles, J. and Varga, L. (2008) Trends in the liquidity of Hungarian financial markets – What does the MNB's new liquidity index show?. *MNB Bulletin* 3(1): 44-51.

Turner, P. (2009) *How have local currency bond markets weathered the financial crisis?*. Script.

http://www.bundesbank.de/download/vfz/konferenzen/20091112_13_frankfurt/paper _turner.pdf.

Note: Zoltán Monostori is also a junior analyst at the Magyar Nemzeti Bank, the Central Bank of Hungary. The views expressed in this paper are those of the author and do not necessarily represent the official position of the CUB or the MNB. The author wishes to acknowledge the help of Csaba Csávás and Zoltán Reppa in commenting on a draft of the paper.

Privatization and market liquidity in Croatia and Serbia

Goran Zafirov¹

¹ School of Business, Stockholm University gz@fek.su.se

Abstract. In 2006 and 2007 Croatia has privatized two of its largest companies by selling part of it to a large number of its citizens, while Serbia has in 2010 and 2011 distributed part of two of its largest companies to its citizens. Listing of these companies onto the stock exchange should lead to increased diversification opportunities for investors, lowering of risk, and thus to the improvement of liquidity in the stock market. Our results show that this effect is largely missing. The spreads have in certain cases narrowed following these privatizations, but have not affected most of the stocks in our sample.

Keywords: voucher privatization, emerging stock market, market liquidity

1. Introduction

During the previous several decades transition economies (former Communist states) as well as developing economies have undergone a process of privatization and development of financial markets. Goals for privatization of companies can be many, and one of them thus is the formation and development of financial markets (Boycko et al, 1994; Jones et al, 1999; Megginson & Netter, 2001; Bortolotti et al, 2007; Dewenter & Malatesta, 1997; Jones et al, 1999; Bekaert & Harvey, 2003). Bortolotti et al (2007) reviewing previous studies state that through a listing the privatization initial public offerings (PIPOs) may give rise to a positive externality and jumpstart an economy's stock market. This occurs with the widening of diversification opportunities present in the market for investors (Bekaert & Harvey, 2003).

Bortolotti et al (2007) study the effect of share issue privatisation (SIP) on market liquidity for 19 OECD countries in the period 1985-2002. They find that market liquidity has improved following an SIP. Their results also indicate that the performance of the stock in the previous six months has an effect on the liquidity of

the market. They focus on the established stock markets which gives us room here to test the effect of SIP on two emerging and highly illiquid markets.

Croatia has privatized and subsequently listed two companies on its Zagreb stock exchange (INA in December 2006 and HT in October 2007). To date, Serbia has also privatized two of the largest companies in the economy and listed them on the exchange (NIS in August 2010 and Belgrade Airport in February 2011), while several more are awaiting the same process. Have the privatizations of these companies improved the stock markets in the respective countries? Has liquidity improved in these markets following these privatizations?

We study the effect that privatization and subsequent listing of these companies had on the liquidity of these markets. Liquidity inherently is a complex concept in the financial markets as it often may give information on some other aspect of the development of the market. The interest in the above stated research question also resides in the fact that these markets are highly illiquid.

Privatizations studied here have implied that the ownership structure of the companies is broadened by providing a large part of the population with the possibility of owning some part of the company. This may lead to increased awareness of the stock market in the population and may lead to spillover of broadening ownership structures to other listed companies. This may lead to the presence of more uninformed traders. More uninformed investors will decrease the problem of adverse selection for institutional investors and thus they are also able to reduce the risk of their investments, and subsequently the costs of trading. Cost of trading in this respect is reflected in the bid-ask spread of any given share. Thus lowering the adverse selection component of the spread, the spread will then (other things constant) narrow, and the market will be more liquid.

As Lesmond (2005) shows high transaction costs may outweigh the information signal for a potential investor, thus he elects not to trade, which in turn may result in observed zero return. This then does not give a potential investor the sufficient diversification opportunities. The market will then be illiquid if the potential investor does not have a timely opportunity to invest into or out of the positions he wants.

We examine the above research question using individual stocks listed on the Zagreb (Croatia) and Belgrade (Serbia) stock exchanges. We try to answer the research question using the smaller event window -3 months before and after the event - as well as using the entire sample period of January 2005 to July 2011 to study the development of market liquidity in these two markets.

In the event period we study simply what happened with the liquidity of each company after the listing of the related company¹, while using the entire sample period we also control for other factors that may influence the level of liquidity in the market other than the effect expected in this study.

Our results of the event study show that some significant results can be found. The privatization of INA has significantly improved the spread ratio, while the privatization of HT has not had the same effect. In Serbia the privatization of NIS has improved the non-benchmark spreads (absolute and relative spread), while it has not improved the benchmarked spread ratio. Privatization of AERO has on the other hand

¹ Related here implies that we study the effect privatization of INA and HT only the Croatian stocks, while the privatization of NIS and Belgrade Airport on the Serbian stocks

lead to the improvement of the spread ratio but has not impacted the nonbenchmarked results.

In the regression model we see significant results for certain stocks. These have not been widespread enough for a conclusion that the effect of privatization on market liquidity actually exists. At times even the coefficients are significant and positive, which means that the liquidity (spread) has deteriorated following these privatizations.

One fact observed in the data is that the turnover has significantly increased following these privatizations, and thus we may note that the result of the privatizations may have been portfolio rebalancing, and may not lead to lowering of risk of investing and thus the spreads may not have reacted positively enough to the privatizations.

Previous studies that have dealt with the development of markets after privatization have rarely studied the effect that it had on market liquidity. Another contribution of the study is that this is the first study to study the effects on market liquidity of mass privatization. Bortolotti et al (2007) study the effect on market liquidity, but they do not use emerging markets, individual stocks nor do they use intra-day information, as they only study price impact on a daily level of the market index. Thus this study tries to fill this gap in the literature.

Another contribution is the policy effect of the results. Since the execution of the privatization was different in Croatia (sold shares) and Serbia (distributed free shares) we may also see if a difference in the results leads to policy effects for future governments examining the issue of privatization. Finally since four more SOEs are awaiting privatization in Serbia the results of this study may also shed light for them how liquidity may change following these privatizations.

Next section of the paper details the literature. Concluding Section 2 are the details of the privatizations and companies involved in Croatia and Serbia. We present the methodology of the study as well as the liquidity measures studied in Section 3, while the data used and the stock selection process is described in Section 4. We finally present the results of the study and the conclusions that the study makes in Sections 5 and 6.

2. Literature Review

2.1. Privatization Theory

Many different reasons for privatization have been put forth. These include the political reasons of government trying to increase voter support for privatization and improve the chances of being re-elected by tying the interest of the population to the interest of the incumbent government (Boycko et al, 1994; Biais & Perotti, 2002; Jones et al, 1999). They look specifically at allocation of underpriced shares and conclude that this is a way for governments to align the interests of median-class voters to the interests of market-oriented governments.

Also privatisation of state owned companies may also be performed in order to boost the funds available to the government for diverse investment projects, or consumption increases.

The design of the privatization program will impact the subsequent performance and success of privatization in a country and the decisions required are dictated by politics (Boycko et al, 1994). One of the decisions in the privatization process as Megginson & Netter (2001) state is a pricing decision and the amount of underpricing and how the price should be formed. Another decision is how to allocate shares, whether one group should be favored to another group. Jones et al (1999) find that governments will consistently underprice SIP offers and will favor domestic investors.

In light of the political reasoning Dewenter & Malatesta (1997) compare the privatization programs in eight developed and emerging markets in order to test whether underpricing is more severe for state privatization companies than for private IPOs. If there are political reasons behind privatization programs then we should see more underpricing for privatized companies than for private initial public offerings (IPOs). They conclude that even though several factors do play a role in the initial returns of the studied companies, in general it cannot be claimed that state IPOs are more underpriced than private IPOs, which is a surprising result.²

This study focuses on the third goal of privatization highlighted by Megginson & Netter (2001), thus we study what effect has privatization of companies had on the market development. Market development here is proxied by market liquidity.

Boycko (1994) as well as Brada (1996) in Megginson & Netter (2001) review several methods that governments can use to privatize SOEs. One of the highlighted methods is share issue privatization (SIP) in which governments sell of a part or all of the government's stake in the SOE to investors in a public share offering. Megginson, et al (2000) in Megginson & Netter (2001) review the choice of SIP versus direct sales of SOEs and find that SIPs are more likely to be used when capital markets are less developed, possibly as a way to develop the capital markets. Jones et al (1999) confirm that this is the method of choice outside of former communist countries.

Another method employed throughout the Central and Eastern European countries is "mass" privatization and its subset voucher privatization. In this method a government distributes vouchers (either free or at some nominal cost) to eligible citizens. These vouchers can be used to bid for stakes in the SOEs.³ Megginson & Netter (2001) write that voucher privatization is the least economically productive privatization method, but in most cases has been used when governments have had few other realistic options, as was the case for Eastern European countries.

There are some differences between these two methods in the goals as well as execution in that the SIP may not involve a large part of the population, while the voucher privatization is performed so that as wide as possible share of the population obtains stakes in the company.

² See Megginson & Netter (2001) for a review of other studies comparing state to privatesector underpricing

³ Design of voucher auctions in which the given vouchers are exchanged for shares is given in Boycko et al (1994, pp.261-265)

2.2. Privatization and Liquidity studies

It may be assumed that these new companies on the stock market will lead to increased net capital inflows into the market. This may occur if foreign investors now see larger diversification benefits in that specific market and decide to enter it. Henry (2000) states several studies that conclude that higher capital inflows will increase the market liquidity. Thus it may be expected that the SIPs may help improve the market liquidity through this channel.

Bortolotti et al (2007) study SIPs in 19 OECD countries for the period 1985-2002. They do not study the emerging markets since in their view SIPs have often coincided with removal of restrictions for foreign investors and thus they may not be able to disentangle the two effects properly. In this study we do not have that problem, since in both countries there have been no significant changes in foreign investment restrictions in the several years prior to these SIP taking place, and thus we will also be able to judge only the effect of SIP on market liquidity.

Bortolotti et al (2007) find that the SIP's in the 19 OECD countries have increased the overall market liquidity and have also found spillover effects to the liquidity of non-privatised firms. As they write, this is an implication of improved risk diversification opportunities and risk sharing brought about by privatization.

Previous studies on the effect of voucher privatization on market liquidity have not been found.

Studied markets have underdeveloped markets. One characteristic of such markets are insufficient risk diversification opportunities for investors. This may lead investors to elect to not take part in the market. Another issue in underdeveloped markets is information transmission and transparency. Privatization and listing of these companies onto the stock exchange may help alleviate these two problems.

Bortolotti et al (2007) state that SIPs may reduce illiquidity by improving the diversification opportunities present. This occurs when the stock markets are trapped in a low liquidity-high risk premium equilibrium. Investors may diversify only when many firms go public. If investors anticipate too few IPOs, they do not enter the market, which remains small and illiquid. Thus a privatization policy which aims to increase the number of IPOs of SOEs may remove this problem and create a sufficient gain of diversification opportunities for potential investors to enter the market.

Jain-Chandra (2002) finds that stock market liberalisations lead to increased liquidity and efficiency in the emerging markets. First, liberalisation will lead to the presence of foreign investors, this increases the investor base in the market, and leads to improved liquidity. Second, the foreign investors will require improved transparency and disclosure rules which will improve the information reaching the market which in turn will increase the efficiency in the market, and help reduce the information asymmetry and subsequently the risk of investing in the market.

Privatization of the studied companies may bring about the increase in transparency in the market if they themselves set new standards which the other will then follow. In Serbia NIS has received an Award from the Belgrade Stock Exchange in 2011 as the company with the best Investor Relations. Thus they may lead the pack in increasing the transparency in the market.

In Croatia both INA and HT are cross-listed on the London Stock Exchange as well. Bortolotti et al (2007) state that cross-listings may enhance foreign investor's

recognition and participation in the privatized stocks. This will lower the overall risk borne by domestic investors. This effect is then coupled with better diversification opportunities in the market and in the end will lead to lower required risk premium. In the end the result of this is, as Bortolotti et al (2007) write, increasing the liquidity of private securities.

2.3. SIPs in Croatia and Serbia

In general in Croatia and Serbia companies have been privatized using direct asset sales (Brada,1996, in Megginson & Netter, 2001) where they have been sold to single or small group of investors. However, the companies studied here have undergone mass privatizations. One of the intents of using a different method in the cases of these companies is to broaden the ownership. A key argument for this method is that it distributes the wealth back to the population.

In Croatia the choice of pricing of the vouchers was that each citizens could buy shares at some nominal cost different for each company and buy a certain (capped) amount of shares. While in this way the government received revenue from selling its shares in the companies, the method in essence discriminated against the lower income part of the population, which may not have been able to buy shares.⁴ On the other hand, since the wealth was effectively distributed against a smaller part of the population, the citizens that have taken part, may have obtained sizable returns both nominally and relative to the original cost of the voucher.

In Serbia, in contrast to Croatia, shares for these privatized companies have been distributed at no cost to each citizen of age willing to take part. This has led to the fact that the government has received no income for its stakes in the privatized companies and there has been no discrimination against lower income citizens. Since nearly entire population has taken part in the program, however, the size of the distributed wealth that each citizen has obtained is nearly negligible.

In 2006 Croatia sold 15% of INA shares to the general public at underpriced value. These shares have then been listed on the 1st of December 2006 on the Zagreb Stock Exchange (ZSE). Each citizen had an opportunity to buy a maximum of 22 shares at 1960 Croatian kuna (HRK; ~265€). According to Kračmer Čalopa et al (2008) INA has seen its stock price increase by 68% in the first hour of trading. During the first day of trading the price opened at 2100 HRK, was highest at 2500 HRK and closed at 2170 HRK (observed in SIRCA database). The importance of the company to the Zagreb stock exchange is undoubted as it alone commands more than 10% of the total market capitalization on the exchange, as of November 2010.

The listing of *Hrvatske Telekomunikacije* (HT, eng. Croatian Telecommunications) has occurred in October 2007, when the Croatian government has sold at underpriced value 25% of the shares to the citizens willing to take part in the privatization. At that time each citizen had an opportunity to buy a share of the company at 265 HRK (\sim 36€). First day of trading was on October 9, 2007. The closing price on that day was 366.5 HRK, which implies a first-day return of nearly 40%. HT commands around 12% of entire market capitalization of the ZSE at the end of 2010.

⁴ For instance, each citizen could buy shares of INA in 2006 at 1960 HRK per share, while the average salary in Croatia at the time was around 4500HRK.

Megginson & Netter (2001) have found similar figures for the first-day-return of SIPs and the market capitalization in other countries, similar to these reported figures for Croatia.

The opening day's returns while significant should also be examined nominally, to see the effect on each citizen taking part in the program. While the amount of 1960 HRK for 22 shares was not directly available to each citizen, a large number of citizens in Croatia have loaned money in order to buy the shares. Thus a citizen which bought 22 shares in INA would receive around 200 HRK per share profit at the close of the first day if the shares were sold on the first day. Thus for 22 shares profit would be around 4400 HRK which is at the level of the average monthly salary in Croatia. Thus the possible profit to be made by each citizen is certainly not negligible.

In Serbia a voucher privatization program has been initiated in 2010. The goal of the program is to privatize and subsequently distribute shares in six large SOEs. As part of the SIP Serbian government is distributing 15% of each of the six large utility and infrastructure companies among its entire population of age, i.e. each citizen is given a certain number of shares of each company.

The first SOE to go through the program in Serbia was NIS (*Naftna Industrija Srbije*; oil company). On the Belgrade Stock Exchange NIS has been listed onto the exchange and the trading with its shares begun on the 30th August 2010. Prior to the listing each Serbian citizen has received five shares of the company, and current and former employees a larger number of shares. The first day's closing price for the shares was 506 RSD (~5€). The market capitalisation of the company alone is more than 8% of the total market capitalisation of the exchange.

A second SOE in Serbia has seen its listing on the Belgrade stock exchange in the beginning of 2011. Each citizen has received one share in the company, which was initially valued at 515 RSD per share. At the end of the first trading day on 7th February 2011, the stock price has reached 632 RSD, a one-day return of 23%. However, subsequently the price started reverting back and is now (October 2011) below the initial value of 515 RSD. The company is smaller than the other SIP companies in this study, since it is responsible for around 1.9% of the market capitalization of BELEX.

The two cases in Serbia highlight the issue that the size of redistribution was fairly negligible to each citizens' wealth. The first day's close of NIS on the exchange was 506 RSD. For five shares thus each citizens could obtain roughly 2500 RSD. At the time the average monthly salary in Serbia was around 34000 RSD. For AERO the first day's closing price was 632 RSD, which is the amount each citizen could obtain since they obtain only one share.

3. Method

The methodology focuses on individual companies in these markets and examines the effects of the listing of the large previous SOEs on the liquidity of each of these stocks. The first methodology is a simple event-time study to check for the effects. After this we proceed to a more general model which should lead to conclusions looking at a longer time scale and controlling for more factors that may impact market liquidity. Using the individual stocks instead of a market index or aggregated data we are able to directly assess whether the privatization programs have led to liquidity spillovers, i.e. whether the companies already listed on the exchanges have seen the liquidity of their shares improve after the listing of the privatized companies.

The first way to check for this effect is to see if liquidity changes in the period surrounding the event, i.e. the listing of the privatized companies. The event window is three months before and after the event, and we use also one month windows for robustness check of the results. We also use the entire sample period in order to see how the liquidity has changed over a longer time period and control for more factors.

We calculate the average level of the liquidity ratio in the period before and in the period after the event. We then take the ratio of after and before, and a value lower than 1 (for spread) would imply improved liquidity since the liquidity ratio is lower after the event than before the event. We perform a simple t-test to check whether the ratio is significantly different from 1.

Increased volatility in the market should, other things equal, decrease the level of liquidity in the market, as it signals increased risk in the market. Thus we add realized volatility of the individual stock. The model that we test is the following:

$$L_{k,t} = \alpha_k + \beta_1 \sigma_{k,t} + \beta_2 D_t + \varepsilon_t \tag{1}$$

Where $L_{k,t}$ is the liquidity ratio of stock k on day t; $\sigma_{k,t} = r_{k,t}^2$, $r_{k,t}$ is the return for stock k on day t; D_t is the dummy variable that takes the value 1 for days after the event (listing of the privatized company) and 0 otherwise. The model is performed for each of the events separately. We also perform a joint model for each country in which we in a single regression test the effect of both privatizations on the liquidity. Thus this model is:

$$L_{k,t} = \alpha_k + \beta_1 \sigma_{k,t} + \beta_2 D_{1,t} + \beta_3 D_{2,t} + \varepsilon_t \tag{2}$$

The idea of using a benchmark market is to use it in order to control for exogenous events. The time period of the study is long enough so that exogenous events are bound to affect market conditions. Thus in this way we attempt to disentangle the effect of privatization on liquidity from other events.

In order to be able to perform this we use a market which in many respect is similar to the studied Croatia and Serbia. For the benchmark market we select the Slovenian Ljubljana stock exchange. Slovenia is chosen since it also is a former Yugoslavian state, thus has had a same starting point in the previous decades. It is similar to Croatian and Serbian markets in size and turnover (see Table 1). Due to its size it may be prone to the same exogenous shocks and should give us the possibility to control for them.

We measure liquidity through the relative bid-ask spread and use it as a dependent variable. In order to control for exogenous shocks we benchmark it to another market's spread. This is done by calculating a ratio of the individual firm's spread to the benchmark market's spread for each day of the sample. Spread is usually considered to be a transaction cost in the market (Bortolotti et al, 2007; Hasbrouck, 2009), and the higher the spread is for a given stock, then the higher transaction cost will be for the average investor, then less trading is expected to occur and the market will be illiquid. Thus measuring the spread may help determine what effect on transaction costs did the SIPs have, and how have the transaction costs in these markets evolved during the recent period.

For each stock the measure is then aggregated into a daily figure in order to be usable in the calculations. Since for each minute we may not have updated information, we use the previous minutes spread for the next minute as well if no update has occurred. This series of spreads across the day is then averaged for each day and each stock.

In order not to have any outlier problems we truncate spreads larger than 25% to 25%.

4. Data

We use daily and intraday data for trades and the order book for two Balkan countries (Croatia and Serbia) as well as for Slovenia. The data obtained is from the SIRCA database. Since the first SIP is that of INA on 1st December 2006 we will use data beginning with 1st January 2005 and onwards. For BELEX we use a different starting point for the spread measure, since the intra-day data on BELEX is kept in SIRCA starting on 22^{nd} June 2006. The end of our studied time period is 31^{st} July 2011.

For Slovenia we have also an issue with the spread data. Bid and ask prices for all but three companies start in mid 2007. In order to be able to use Slovenia as the benchmark market we calculate the spreads for Slovenia based on end-of-day data.

Table 1 below shows the summary statistics for the three markets. We can see from the table that the number of issues on the Belgrade stock exchange is much larger than on the other two markets. The reason for this is that in Serbia a large number of privatized companies had an obligation to list their shares on the stock exchange following privatization. This then in practice implies that a large number of these stocks are very rarely traded. Another notable fact is that Belgrade and Ljubljana stock exchanges are similar in size and turnover, while Zagreb and Ljubljana stock exchange had a similar yearly share turnover.

	Zagreb	Belgrade	Ljubljana
Number of issues	254	1410	70
Market capitalization (€ bill.)	27.02	9.53	7.03
Turnover (€ mill.)	7,90	2,02	2,07
Yearly Turnover (2010)	6,37%	2,13%	7,01%
mean return	0,032%	-0,078%	-0,018%
mean return pre crisis	0,094%	-0,038%	0,087%
mean return crisis	-0,065%	-0,11%	-0,10%
standard deviation	1,58	1,88	1,31

Table 1 – Summary statistics for the studied markets.

The number of shares listed and the number of stocks traded in these markets forces us to think about which stocks actually to use in our calculations. Information for each stock may not be available continuously or may not be accurate if no trading or quote updating has occurred in a number of trading days. Thus, we have to establish some criteria that a stock needs to satisfy in order to be included in our final sample.

The criteria that we use is based on the number of trading days of each stocks. For each stock market we look at the period starting with one year before the first privatization.⁵ For the entire period after this start we select only the stocks that have at least 75% of the trading days with non-zero turnover.⁶ This criteria is selected in order to minimize the possibility of stale quotes and information. On the other hand we also attempt to maximize the stock sample as well, and for this reason we do not select a higher criteria.⁷

Figure 1 shows the ratio of the Croatian/Serbian value-weighted relative spread to the Slovenian value-weighted relative spread.

5. Analysis

The results of the event time methodology are not shown here for brevity purposes. We simply calculate the average in the three months before a privatization and three months after the privatization. The ratio of after-to-before then gives us information of whether the spreads have decreased (ratio lower than 1) or increased (ratio greater than 1), and use a t-test to check whether the ratio is significantly different from 1 and in the basic form test the effect of the privatization on the liquidity of the market. We first test whether the absolute spread (nominal for the individual company only) ratio of after-to-before is different from 1. The t-test shows that that the absolute spread has not changed following the privatization of INA. Similarly the relative spread has not been influenced by the privatization of INA. On the other hand, the relative spread ratio of Croatian companies to the Slovenian market, has been influenced by the privatization of INA. This may show that some exogenous event has impacted the Slovenian market, thus increasing its spreads, while the spreads in Croatia have remained largely the same. Another explanation is that the privatization of INA has countered the exogenous events and their impact on liquidity, and in this respect improved the liquidity (spreads) in the Croatian market.

For the privatization of HT a similar picture is seen for absolute and relative spread as for the privatization of INA. On the other hand, the privatization of HT has not even improved the spread ratio.

In contrast to the Croatian privatizations, NIS privatization has brought about a significant decrease in the absolute and relative spread in the Serbian market, while the spread ratio to the Slovenian market has not been impacted by the privatization of NIS. An explanation for this discrepancy is that some exogenous events may have

⁵ For Croatia the period starts in 1st December 2005 as well as for Slovenia, while for Belgrade this period starts on 30th August 2009.

⁶ CROBEX as part of its rules has set up different classes of liquidity to which each stock is classified, based on the percent of days traded and market depth. The first class of stocks entails that each stock should trade at least 75% of trading days. This may be an argument for selecting the above set criteria.

⁷ Summary statistics for each individual stock are omitted for space requirements, and can be obtained from the author

impacted the spreads in Serbia, while the privatization of NIS may have only a small impact and thus has not been significant.

A stark contrast can be seen for the privatization of AERO. In this case the absolute and relative spreads have not been impacted by this privatization. On the other hand, a spread ratio has been highly impacted by this privatization. For all ten companies listed on the market the spread ratio is lower than 1.

Table 2 shows the results from our regression in equations 1 and 2. The first two columns in the table show the coefficients for volatility and the privatization dummy of the first privatization, and in bold are given values significant at 5% significance level. The results show that INA has had an impact on the spread ratio in most of the companies in the sample. However, five of the 26 significant coefficients are positive, implying widening of the spreads and thus lower liquidity. Similar result is seen for the privatization of HT (columns 3 & 4). Finally in the joint regression (columns 5-7), where we together test for the impact of INA and HT on the spread ratios we observe that less coefficients are significant, and that more coefficients after the privatization of HT are now positive than before.

Turning to the results from Serbia, we see that significant results (column 2) are obtained for 5 of the 10 companies, of which one is with a positive coefficient implying that the spread has widened following the privatization of NIS. The privatization of AERO has resulted in significant results for 7 of the 10 companies, and in all cases the coefficient is negative implying that the spread has narrowed and that liquidity has improved for these companies. In the joint regression significant results are scarce, as both privatizations have lead to only 3 each significant changes in the spreads, one of them with positive coefficient.

6. Conclusion

The markets studied here are underdeveloped and illiquid by comparison to any other developed market. The changes in these markets while observable are still far from creating a well-functioning market as characteristic for developed economies. Tying with the topic studied in this study Megginson & Netter (2001) state that the countries which emphasized asset sales or voucher privatizations (as has occurred in Croatia and Serbia) seem to lag behind in market development.

The results in this study show that mass privatizations in Croatia and Serbia have had some significant results, but these are far from systematic enough in order to draw a conclusion that the privatization has improved liquidity in these markets.

While we have found some results that confirm a significant effect of privatization on market liquidity, it is hard to disentangle the actual channel and process through which the market liquidity is affected. While we go from the assumption that increased diversification benefits in the market and increased presence of the population in the market will lead to lower risk and required return, these assumptions we have not been able to test directly. Thus a further study may be necessary in order to test for the actual channel through which liquidity is affected.

We have observed that turnover has increased in the two markets following the privatizations, while spreads (liquidity) appear not to follow this development. This may then imply that the increased turnover may have lead to rebalancing of portfolios, thus increased turnovers, but the risk and subsequently spread and liquidity, may not have been affected sufficiently.

References

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5 (1), 31-56.

Bekaert, G., Harvey, C.R., 2003. Emerging markets finance. *Journal of Empirical Finance*, 10(1-2), 3-56.

Benić, V., Franić, I., 2008. Stock Market Liquidity: Comparative Analysis of Croatian and Regional Markets. *Financial Theory and Practice*, **32**(4), 477-498.

Biais, B., Perotti, E., 2002. Machiavellian Privatization. *The American Economic Review*, Vol.92, No.1, pp.240-258

Bortolotti, B., de Jong, F., Nicodano, G., Schindele, I., 2004. Privatization and Stock Market Liquidity. *Journal of Banking & Finance*, **31**(2), pp. 297-316.

Boycko, M., Shleifer, A. and Vishny, R.W., 1994. Voucher privatization. *Journal* of Financial Economics, **35**(2), pp. 249-266

Dewenter, K.L., Malatesta, P.H., 1997. Public Offerings of State-Owned and Privately-Owned Enterprises: An InternationalComparison. *The Journal of Finance*, Vol.52, No.4, 1659-1679

Hasbrouck, J., 2009. Trading Costs and Returns for U.S. Equities: Estimating Effective Costs from Daily Data. *The Journal of Finance*, Vol.64, No.3, pp.1445-1477

Henry, P.B., 2000. Stock market liberalization, economic reform, and emerging market equity prices. *Journal of Finance*, Vol.55, 529–564.

Jain-Chandra, S., 2002. *The Impact of Stock Market Liberalization on Liquidity* and *Efficiency in Emerging Stock Markets*. Working Paper edn. SSRN

Jones, S.L., Megginson, W.L., Nash, R.C., Netter, J.M., 1999. Share issue privatizations as financial means to political and economic ends. *Journal of Financial Economics*, Vol.53, pp.217-253

Lesmond, D.A., 2005. Liquidity of emerging markets. *Journal of Financial Economics*, **77**(2), 411-452.

Megginson, W.L., Netter, J.M., 2001. From State to Market: A Survey of Empirical Studies on Privatization. *Journal of Economic Literature*, Vol.39, No.2, pp.321-389

Minović, J.Z., Živković, B.R., 2010. Open Issues In Testing Liquidity In Frontier Financial Markets: The Case Of Serbia. *Economic Annals*, Vol.55, No.185, pp.33-62

Ritter, J.R., 1991. The Long-Run Performance of Initial Public Offerings. *The Journal of Finance*, Vol.46, No.1, pp.3-27

Law on the Privatization of INA Petroleum Industry [in Croatian] (2002) Narodne novine [Official Gazette], 32/02.

TERRORISM AND CAPITAL MARKETS: THE EFFECTS OF THE ISTANBUL BOMBINGS

Nikos Christofis¹, Christos Kollias², Stephanos Papadamou³, Apostolos Stagiannis⁴

¹Leiden University, The Netherlands, n.christofis@gmail.com

²University of Thessaly, Department of Economics, kollias@uth.gr

³ University of Thessaly, Department of Economics, stpapada@uth.gr ⁴ University of Thessaly, Department of Economics, astagiannis@uth.gr

Abstract: Beyond the loss of life and personal injuries that the victims of terrorist actions suffer and the atmosphere of fear terrorists seek to create with their premeditated use of brutal violence, terror also has real economic and political costs that go beyond the immediate costs and damages of a terrorist attack. Terrorist actions can have a multitude of economic consequences that may adversely affect a number of economic indices, sectors and activities including their impact on capital markets. This paper sets out to examine the interpretation of the terrorist attacks, the political interactions and the effect of these attacks on Turkish Stock Market. It focuses its empirical investigation on the Istanbul stock-market and the impact that recent major terrorist incidents exerted on market behaviour. Findings reported herein indicate in some cases significant abnormal returns

Keywords: terrorism, capital markets, conditional volatility

1 Introduction – Turkey in the new era

Turkey is one of the few exceptions in the Middle East of a country that is undergoing such a rapid globalization in all fields, even compared to other countries' progress of which their "*europeanness*" was never disputed. That procedure placed her into one of the countries of the global capitalist system. The role Turkey acquired as a mediator between the "East" and the "West" indicates that "the perspective of new regionalism better reflects the actual regionalist projects in the contemporary era" and that, "since it puts the emphasis on the conceptualization of new regionalism as a means of riding on economic globalization, it often overlooks the aspect of globalization as *facilitator* of the new regionalism" (Onder, 2008: 87-88).¹ This is true for Turkey, especially considering her increased integration into the world capitalist system through the neoliberal reforms in the 1980s under Turgut Özal (Acar, 2002; Öniş, 1998; Öniş, 1991). In addition, the leading role in international institutions, such as the International Monetary Fund (IMF), the World Bank and World Trade Organization (WTO) helped tremendously towards that direction.

Despite the economic progress however, it is often argued that it is the countries of the semi-periphery (Ozdogan, 2006) and not the centre that are hit by terrorist attacks, although the 9/11 attacks in the US, as well as the 7/7 attacks in London, should make us skeptical about the validity of such a conclusion (Crenshaw, 1981). This is true for Turkey of which the leading regional role in the Middle East put her on the target list of international terrorism. Some of them, Turkey being just one of them, indicate that they are on this list for quite some time (Mango, 2005), mainly because of its choice to ally with the western hegemonic powers since the 1950s and most notably the United States of America, following the westernizing/modernizing paradigm instigated by the founder of Turkey, Mustafa Kemal (Atatürk).

The formula centre – periphery – semi-periphery formulated by the dependency and world system scholars is quite revealing in that respect. As the formula reveals the world is divided into three categories: core, periphery and semi-periphery. According to that theory the world is a unified social and economic system dominated by the core countries, i.e. the developed ones, while the countries of the periphery, i.e. the underdeveloped ones, are used by the former to establish industry where labor and raw material are cheaper than they should be. However, the new social, political and economic relations produced by the industrialization and urbanization of the new world-system order are characterized by less social cohesion that promote inequalities against the core countries, while at the same time pathological/complex conditions, as Fanon (Fanon, 2008) argued,² are produced in the

¹ Two major dynamics explain the Turkish state's pursuance of new regionalist projects: 1) the end of the cold war and the collapse of the Soviet system and 2) neoliberal economic globalization.

² Similar lines following Fanon's paradigm can be drawn replacing colonialism with globalization and its effect upon underdeveloped countries that wish to imitate other countries' globalization process, in order to appropriate and imitate the cultural, political and economic code of the '*globalizer*' (colonizer).

underdeveloped countries (Wallerstein, 2001; Mardin, 1973). Furthermore, as it becomes obvious the political and economic order is more likely to be disordered, because of the vulnerability of the new status, while, adding to that, shaky 'democratic' institutions, or the lack of, or little, democracy leads to political violence, and in extent, it can be argued to terrorism (Porta, 1995; Tarrow, 1994).

Semi-periphery states fall into this scheme as a buffer zone, between the core and the periphery. These states prevent direct confrontation and conflict of the states mentioned. Turkey is one of the semi-periphery countries that are striving to prevent direct conflict between the two. Thus, schematically and rather simplistically, we could use the example of the triangle, U.S.-Turkey-Iraq. The U.S., being the core country, attempted to make use of the military bases of a semi-periphery country, i.e. Turkey, to attack a periphery country, i.e. Iraq. This rather simple, but quite explanatory example may help us understand how the world-system works. However, in that scheme it should be noted the difficult position the semi-periphery country can be. If that country, in our case Turkey, fails to act as a 'preventer' of direct conflict and allow the usage of her military bases it will face the increased likelihood of being a target of future terrorist attacks by the periphery countries, while if she does not allow it she will face most probably heavy economic consequences by the core countries.

The terrorist attacks understudied should be seen through that prism rather the one that very often looked through and was propagated by academics such as Bernard Lewis (1990) and Samuel P. Huntington (1993). Their approach is that in fact, religious extremism, and in extent religious terrorism, replaced ideology, and to be precise communism, with a cultural and religious divide. This view has been subjected to intense criticism by a number of scholars, but its influence is still dominant among scholars and, with much more obvious catastrophic effects, in the media. The danger of this approach is "to overlook the fact that militant jihad movements and terrorism are not just the products of warped individuals or religious doctrine, whether mainstream or extremist interpretations, but of political and economic conditions" (Esposito, 2002: 152). It seems however, that the U.S., especially after 9/11 terrorist attacks is inclined to impose its policy upon the world based on that false theory, and moreover, make the world to believe that such a theory does have a basis. However, what the U.S. managed to do is to actually split the world into two zones, the war zone and the peace zone (Aras & Toktas, 2007), by replacing the old Cold-war enemy, communism, with another constructed 'enemy', Islam/religious fundamentalism, in order to legitimize her own agenda, or actions. As it was presented by the US themselves, that was the "democratization" of the Middle East (Albrecht and Schlumberger, 2004),³ according to the western values, as if Islam comes in direct contravention with modernism.

³ It is argued that "authors" mistakenly "examine what did not exist, instead of what was actually going on in the Arab world". The fact that democracy is a daily plebiscite, to paraphrase slightly Renan, and nothing indicates that such a process occurred in the Middle East during the last decade (not that it is impossible) should shift our attention not to the 'failure' of democratization, but rather to the 'success' of authoritarianism.

2 The Istanbul Attacks: Who, Where, and How?

Despite the fact that an appropriate conceptualization of terrorism would save us from talking abstractly, or rather too broadly, incorporating and excluding most of the times factors that are mandatory to terrorist analysis, scholars disagree, as it is the common with all *-isms*. Similarly, we will choose a rather broad description of terrorism, being that *the use, or the threat of violence in order to make known the motives and political goals of the group(s) that use violence, and hence, to fulfill those motives and goals forcing the targeted countries to comply with their demands. Additionally, our description is referring <i>only* to non-state actors, and therefore, cannot claim any further validity besides the one referred to.

We will limit our case on the Istanbul terrorist attacks during the past decade, and specifically, on those of 1999, 2003, and 2008. Of these three, the one in 2003 is considered as the most horrific from the point of view of, both, human casualties, being of course the most important in such occasions, and material damages, let alone the distress and psychological suffering caused to the people.

On 13 March 1999, a terrorist attack took place in the Mavi Çarşı shopping mall in a suburb called Göztepe in the Asian side of Istanbul. The six-story building, where the shopping mall was lodged, was occupied by a woman and four men carrying guns according to eyewitnesses (Cumhuriyet, 14 March 1999). The people were forced to move to the upper floors of the building, while the attackers poured petrol in the entrance of the building right before they left at 16.30h local time, when they threw two Molotov cocktails that set the building ablaze. The attack left 13 people dead, 3 burned alive while 10 died from the heavy fumes and 23 people were injured (Rodoplu, Arnold, Ersoy, 2003; Hürriyet, 14 March 1999). The following day another terrorist attempt took place at the Burger King in the Istanbul region Avcılar, but the bomb was luckily diffused with success (Cumhuriyet, Hürriyet, Milliyet and Radikal).

Despite the economic consequences, the attacks had also political implications. The attacks were associated with the *Kurdistan Workers' Party (Partiya Karkerên Kurdistan*, PKK), despite the fact that no-one has claimed responsibility for the attacks. The Turkish media rushed to accuse PKK (Radikal, 14 March 1999), saying that the group was taking revenge for the arrest of its leader, Abdullah Öcalan (BBC, 14 March 1999), who was arrested a month ago in Kenya. PKK's actual involvement is, as it seems, of little importance, since the 'verdict' had already been made, especially after its link with two other extremist groups was established.⁴ Prime Minister Bülent Ecevit expressed the feeling that "they [terrorists] are trying to cause panic during an election period" (Kınıklıoğlu, 2002), which was set for April of the same year. Indeed, the bombings caused disorder to the political life of Turkey, which almost resulted to the postponement of the eminent elections (all newspapers, 24 March 1999). The attacks also caused the ban of pro-Kurdish Med TV broadcasting

⁴ The two other groups are the Revolutionary People's Liberation Party/Front (DHKP/C), coming from Dev-Sol, a leftist guerrilla group of the 1970s (Revolutionary Left), and the Turkish Workers and Peasants Liberation Army (TIKKO). Both of them are associated with PKK.

from England and Belgium for spreading inflammatory statements encouraging acts of violence in Turkey and elsewhere.

The Turkish counter-terrorist extensive measures, already started after the Öcalan's arrest, were increased and they seem to be legitimized after the 9/11 attacks, finding in the U.S. an ally to their new war on terrorism. However, even these measures, not even the emergence of AKP (*Adalet ve Kalkınma Partisi*), a moderate Islamist pro-European party, could not take Turkey off the terrorist list.

The hits in 2003 were the more devastating in the recent Turkish history. They took place with a five-day interval between them. The first ones, aimed at the Jewish synagogues *Bet Israel*, in Şişli, and *Neve Shalom*, in the Galata district, on November 15, when two cars carrying bombs crashed into the synagogues almost simultaneously. Five days later, 20 November, two more trucks exploded. This time the target was the British Consulate, in Galata region and the HSBC Bank, located at the Levent region.

A radical Turkish group, called *Great Eastern Islamic Raiders' Front (İslami Büyükdoğu Akıncılar Cephesi* – İBDA/C) claimed the responsibility of the attack (Faraç, 2003; Karmon, 2003) while there is substantial evidence that this group is closely affiliated with Al-Qaida. The Turkish authorities after investigating the issue claimed that the decision to bomb targets in Turkey had been taken in Afghanistan (Mango, 2005; 68). The attacks that were named "the Turkish 9/11" (Cumhuriyet, 17 November 2003) were so powerful that smashed the 1.5 tone door of the Consulate while a huge cloud of smoke covered the whole centre of the city, and cost the lives of 60 people, including the British consul general Roger Short, while more than 750 were injured (Tavıloğlu, et. al, 2005).

Within the binary opposition, namely the peace zone and the war zone, we referred to before, both of the attacks make sense. The attacks aimed at hitting neocolonialism, or the hegemony of the capitalist core and its allies. However, there is an added importance to these attacks since, al-Qaida, as well as other radical Islamist groups, Turkish or other, consider Turkey's history as a deviation from the right path of Islam. Al-Qaida is a worrying issue for Turkey since, according to Osama bin-Laden "the failure and dissolution of the Ottoman Empire represented the ultimate humiliation of Islamic civilization" (Aras & Toktaş, 2007). Moreover, Turkey's relations with the US, on military and political level, as well as Israel's recognition, signing security agreements, and joint military tasks, put Turkey in the hotspot of international terrorism. It is however, difficult to be sure how these Turkish radical Islamist cells came in contact with al-Qaida and prepared the attacks. Al-Qaida is "transnational in its identity and recruitment and global in its ideology, strategy, targets, network of organizations and economic transactions" (Esposito, 2002; Aras & Toktas, 2007), a fact that makes it even more difficult to calculate a probable future attack, as well as the extent of its communication network and its relations with various individuals or groups in other countries. Hence, the breaking down of any democratic freedom, as it was implied (Mango, 2005), in order to bring significant results against terrorism is highly unlikely.

After a relevant calm period Istanbul was struck again by another attack, this time in the Güngören neighborhood in the European bank of the city, on 27 July 2008. The terrorist attacks were not a suicide attack because as it seems they were triggered by a remote device. The attackers placed first a bomb in a trash can, causing a minor

explosion, to attract onlookers to the scene, while a second bomb was triggered ten minutes later using RDX, an explosive element widely used in military and industrial applications, that cost the lives of 17 people and the injury of to more than 150.

The attacks in Güngören resembled previous attacks by PKK. The explosive element, RDX, was identified as an explosive that was used before in PKK attacks, as the hit at residential areas bore the mark of PKK and in addition, it was believed that the allegedly PKK hit was a retaliation for the 12 Kurdish targets that were hit in Northern Iraq by Turkish jets. While the Turkish authorities were striving to confirm their suspicions about PKK, PKK leader Zubeyir Ayda denied any involvement in the deadly bombings (Today's Zaman, 5 August 2008). However, what complicates the situation more is the Germany's Federal Intelligence Service statement (BND). According to BND Chief Ernst Uhrlau, the attack in Güngören was the work of either al-Qaeda or Turkey's "deep state".

Despite the obvious fact that the statement could be used as to avoid any friction to the AKP rule, since Turkey's accession to the EU seems to be being obtained at a steady pace, it also raises suspicions. AKP already in power for the second time since 2007, is considered, in general, to be trying to democratize and Europeanize Turkey more. Thus, it started working on a new more democratic constitution in June 2007, and was presented in September 2007. However, it faced strong opposition by the secular circles because of AKP's alleged "hidden intentions" (Özbudun and Gençkaya, 2009: 105). The traditional secular elite and the military, the bastions of the Turkish democracy, started, in spring 2007, to create "political maneuverings of dubious legal validity in order to "save the last citadel of the secular republic from the occupation of an alleged "Islamist" (ibid: 97). Secularism, as it was evolved in the Turkish context provides the essence of the "state ideology", and AKP's rule is seen as the erosion of that ideology.

Added to that, AKP launched a case against Ergenekon, an ultra-nationalist organization, having links with military and security forces and was accused of terrorism. It is beyond the scope of our purpose to refer to AKP's measures against Ergenekon, but it is suffice to say that since the establishment of the Turkish Republic, the military was preserving the integrity, unity and modernity of the country, and by the 1990s its political autonomy reached its apogee and managed "to go above and beyond the constitutional authority of democratically elected governments, [and] can also include not only direct but also indirect influences on the government" (Sakallıoğlu, 1997: 153).

3 The impact on Istanbul Stock Exchange

Severe political events such as wars and unexpected terrorist attacks have an impact on stock markets and affect asset prices. The economic effects of terrorism and in particular the impact it has on markets has attracted a considerable and growing body of literature (Bruck, 2007; Bruck, 2005). Building on these studies, at this point the paper examines the impact three major terrorist attacks in Turkey had on the Istanbul Stock Exchange.

Istanbul Stock Exchange has a market capitalization of \$118bn in 2008 and 317 listed companies. For the purposes of the analysis that follows, daily prices of the

Istanbul Stock Exchange (ISE) are used. Six major indices are selected as the most representative sample, the National-100, Bank Index, Industrial Index, Services Index, Tourism Index and the Trade Index. All of them are drawn from the *Reuters DataStream database*. The sample covers the period from January 1997 to December 2009 and includes 3391 trading days (Graph 1). The dates of interest for the purposes of the paper are 14/03/1999, 20/11/2003 and 27/07/2008 when the three selected attacks in Turkey took place.

<Graph 1>

To start with, by applying an event study methodology we try to investigate the impact of the events on the above indices. We attempt to identify differences in the persistence and magnitude of these effects across these indices taking into account the global financial environment and the USD/TL.

Event Study Methodology

By using an asset pricing model - market model according to McKinlay (1997) - in an international framework, the daily return of the Dow Jones index (R_t^{US}) is used, for capturing the effect from the worldwide financial interactions. In the presence of exchange risk, the daily rate of change of the US dollar versus the Turkish Lira currency (FX_t) is added in order to reflect the covariance of the asset with the US dollar exchange rate.

$$R_{t} = a + bR_{t}^{US} + \gamma FX_{t} + e_{t}$$
(1)

Then we estimate the parameters of the regression equation (1), using 3191 trading days from February 1997 to December 1999. Based on equation (1), the expected index return should be as follows:

$$E(R_t) = a + bE(R_t^{US}) + \gamma E(FX_t)$$
(2)

Daily excess returns were measured by the difference of the actual return minus the expected return.

$$AR_t = R_t - E(R_t)$$
(3)

Initially, the event-day abnormal returns are calculated. The date of the event is set at t=0, and two longer event windows are examined by computing the cumulative average abnormal returns (CARs) 3 days following the event (t=3) and 6 days following the event (t=6). The cumulative abnormal returns (CARs) were estimated using the following equation:

$$CAR_t = \sum_{t=T_1}^{T_2} AR_t$$

where T_1 is the event day and T_2 is consequently 3 and 6 days after the event.

The findings of the event study methodology for each one of the three key dates that shaped the whole event are presented in Table 1 where the abnormal returns and statistical significance levels for the 0, 3, and 6-day event windows are shown. Additionally, a column was included showing the number of trading days elapsed before each of the three indices returned to its previous level.

According to the results, the 1999 attack has negative effect on the returns of all indices except from Banks, Trade and Industrial. Regarding the abnormal returns the largest negative and statistical significant abnormal return is on Tourism Industry (-8.79%) and on Industrial Index (-2.65%). Tourism Index displays also significant negative 3 and 11-day cumulative abnormal returns enhancing previous evidence about the sensitivity of tourism industry on terrorist events (Drakos & Kutan, 2003). All other indices present a quick recovery of two days and positive CARs. In this overall good reaction with the exception of Tourism Index may help that the event took place on Sunday and investors discounted the causes and the implications.

The 2003 blasts were devastating for the ISE. On the event day the actual reaction of the main share index was -7.4% and the stock exchange closed and reopened on December 1st with Turkey's Central Bank prepared to intervene to support the falling of Turkish Lira. Without exception, all indices fall, Banks and Tourism Index fall 11% with negative and statistically significant ARs of -10%. Similarly to the previous attack, Tourism Index continued to record losses and rebounded only after the pass of fifteen trading days. The days ISE remained closed proved enough for the indices and the investors to recover in a single trading day. Looking on Graph 2 increased volatility in returns of National-100 Index were appeared after the event.

Our results show no significant negative reaction of the ISE after the July 2008 attacks. Extremely positive and statistical significant CARs for most of the indices are presented in Table 1. Trade Index was the most conservative with 1.62% 3-day CAR. Turkish state's decisions to attribute the double bombing to "PKK" were both predictable and unconvincing. The reaction of equity investors was a testament of the relative insignificance of the incident.

<Table 1> <Graph 2>

5 Concluding remarks

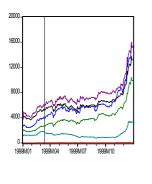
This paper, examine the environment in which the three terrorist attacks took place and the reaction of the Istanbul Stock Exchange. Turkey's strategic geographical position offers advantages and disadvantages. The economic development of the last decade and the leading regional role in the Middle East put Turkey on the target list of international terrorism. The U.S. role in Turkey, the geographical position, the religion and the European perspective of Turkey create a hazy picture that potential disruptors, the "deep state" and speculators could exploit. It is clear that the introduction of transnational terrorism in Turkey change the political and economical rules. Regarding the role of the Stock Exchange in this environment our empirical findings indicate the significance of the 2003 attack and the sensitivity of Tourism Stocks. However, investors show trust in Turkey's high ratios of development and the rebound of the Stock Exchange was very quick despite the severity of the events.

References

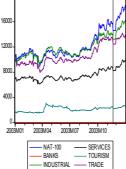
- 1. Acar, F. (2002), "Turgut Özal: Pious Agent of Liberal Transformation" in Heper, M. and Sayari, S. (eds.), *Political Leaders and Democracy in Turkey*, Lexington Books, Lanham, Maryland, 163-180
- 2. Albrecht, H. and Schlumberger, O. (2004), "Waiting for Godot: Regime Change without Democratization in the Middle East", *International Political Science Review/Revue Internationale de science politique*, 25:4, 371-392
- 3. Aras, B. & Toktaş, Ş. (2007) "Al-Qaida, 'War on Terror' and Turkey", *Third World Quarterly*, 28:5, 1033-1050
- 4. Bruck, T. (2007), *The Economic Analysis of Terror*, Routledge, London and New York
- 5. Bruck, T. (2005), "An economic analysis of security policies", *Defence and Peace Economics*, 16:1, 375-389
- 6. Crenshaw M. (1981), "The Causes of Terrorism", *Comparative Politics*, 13:4, 379-399
- della Porta, D. (1995), Social Movements, Political Violence and the State: A Comparative Analysis of Italy and Germany, Cambridge University Press, New York
- Drakos, K., & Kutan, A. L. (2003) "Regional effects of terrorism on tourism in three Mediterranean countries", *Journal of Conflict Resolution*, 47:5, 621-641
- 9. Esposito, J. (2002), Unholy War. Terror in the Name of Islam, Oxford University Press, New York
- 10. Fanon, F. (2008), Black Skin, White Masks, Grove Press, New York
- 11. Huntington, S. P. (1993), "The Clash of Civilizations?", Foreign Affairs, 72:3, 22-49
- 12. Karmon, E. (2003), "The Bombing of Synagogues in Istanbul: An al-Qaida Operation or a Revival of Turkish Islamist Terrorism?", at *International Institute for Counter-Terrorism*
- Kınıklıoğlu, S. (2002), "The Democratic Left Party: Kapıkulu Politics Par Excellence" in Rubin, B. and Heper, M. (eds.), Political Parties in Turkey, Routledge, London and New York, 4-24
- 14. Lewis, B. (1990), "The Roots of Muslim Rage", The Atlantic Monthly
- 15. MacKinley, A. C. (1997), "Event Studies in Economics and Finance", *Journal of Economic Literature*, XXXV, 13-39
- 16. Mango, A. (2005), *Turkey and the War on Terror. For Forty Years We Fought Alone*, Routledge, London and New York
- 17. Mardin, Ş. (1973), "Center-Periphery Relations: A Key to Turkish Politics?", *Daedalus*, 102:1, 169-190
- Onder, N. (2008), "The Turkish Project of Globalization and the New Regionalism", *Alternatives: Turkish Journal of International Relations*, 7:2-3, 86-110
- 19. Ozdogan, A. (2006), "Where do Terrorists Come From?" in Orttung, R. W. and Makarychev, A. (eds.), *National Counter-Terrorism Strategies. Legal, Institutional, and Public Policy Dimensions in the US, UK, France, Turkey and Russia*, IOS Press in cooperation with NATO Public Diplomacy Division, The Netherlands, 21-42

- 20. Öniş, Z. (1998), State and Market: The Political Economy of Turkey in Comparative Perspective, Boğaziçi University Press, Istanbul
- 21. Öniş, Z. (1991), "The Political Economy of Turkey in the 1980s: The Anatomy of Unorthodox Liberalism" in Heper, M. (ed.), *The Strong State and Economic Interest Groups. The Post-1980 Turkish Experience*, Walter de Gruyter, New York and London
- 22. Rodoplu, U., Arnold, J. and Ersoy, G. (2003) "Terrorism in Turkey: Implications for Emergency Management", *Prehospital and Disaster Medicine*, 18:2, 152–160
- 23. Sakallıoğlu, U.C. (1997), "The Anatomy of the Turkish Military's Political Autonomy", *Comparative Politics*, 29:2, 151-166
- 24. Tarrow, S. (1994), *Power in Movement: Collective Action, Social Movements and Politics*, Cambridge University Press, New York
- 25. Tavıloğlu, K., et. al (2005), "2003 terrorist bombings in Istanbul", *International Journal of Disaster Medicine*, 3:1-4, 45–49
- Wallerstein, I. (2001), "The Rise and Future Demise of World Capitalist System" in Wallerstein, I., *The End of the World As We Know It: Social Science for the Twenty-First Century*, University of Minnesota Press, Minneapolis, 192-201

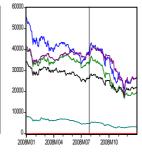
Appendix



Graph 1: National-100, Banks, Industrial, Services, Tourism & Trade Indices

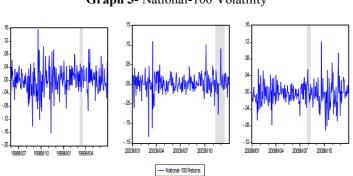


20000



	15/3/1999			20/11/2003		27/7/2008			
Index	Event-day AR	3-day CAR	6-day CAR	Event-day AR	3-day CAR	6-day CAR	Event-day AR	3-day CAR	6-day CAR
Nat-100	-2,27%	1,96%	0,61%	-7,51%	2,05%	4,56%	2,09%	11,19%	10,25%
t-stat	(-0,73)	(+0,08)	(+0,04)	(-2,87)*	(+0,78)	(+1,73)	(+5,02)*	(+5,02)*	(+4,60)*
Days to Rebound	2			1			0		
Bank Index	-1,95%	2,15%	-0,45%	-10,40%	4,30%	7,90%	3,95%	20,57%	18,44%
t-stat	(-0,45)	(-0,01)	(+0,44)	(-2,72)*	(+1,12)	(+2,06)	(+1,06)	(+5,55)*	(+4,98)*
Days to Rebound	3			1			0		
Industrial Index	-2,65%	0,99%	7,05%	-5,97%	2,30%	4,95%	0,56%	6,82%	7,63%
t-stat	(-1,13)	(+0,42)	(+3,02)*	(-2,92)*	(+1,12)	(+2,42)*	(+0,31)	(+3,78)*	(+4,23)*
Days to Rebound	2			1			0		
Services Index	-0,89%	-0,82%	2,60%	-5,97%	0,25%	-0,21%	0,66%	6,34%	4,29%
t-stat	(-0,36)	(-0,33)	(+1,05)	(-3,05)*	(+0,12)	(-0,10)	(+0,39)	(+3,72)*	(+2,51)*
Days to Rebound	0			1			0		
Tourism Index	-8,79%	-16,80%	-14,44%	-10,67%	-10,97%	-8,60%	0,17%	5,10%	14,41%
t-stat	(-2,37)*	(-4,53)*	(-3,89)*	(-4,37)*	(-4,49)*	(-3,52)*	(+0,08)	(+2,49)*	(+7,04)*
Days to Rebound	>100			15			0		
Trade Index	0,08%	-0,26%	3,06%	-6,89%	-1,45%	-1,37%	-0,01%	1,62%	4,69%
t-stat	(+0,03)	(-0,09)	(+1,08)	(-3,71)*	(-0,78)	(-0,73)	(-0,05)	(+0,95)	(+2,76)*
Days to Rebound	0			2			0		

Notes: T-statistics are in parentheses. "Number of trading days for the market index to return to pre-attack level. "Statistically significant at the 0.05 level



Graph 3- National-100 Volatility

Investigating the causality between Financial Development and Economic Growth in the emerging economies of Europe, specifically: Albania and Turkey

Esida Gila Gourgoura

Business Administration and Economics CITY College - International Faculty of Sheffield University, egila@city.academic.gr

Abstract

This paper is a research proposal on "The causality between Financial and Economic Growth in Albania and Turkey". Beyond simply providing a summary of the theoretical considerations on the "finance-growth" nexus and discussing the international empirical work devoted to the issue, the paper is particularly focused on evidence from the emerging economies. It is concluded that further research should be undertaken for these countries, in order to add to their existing insufficient academic literature. Therefore, it is proposed that an empirical investigation should be performed for Albania and Turkey, since they represent two developing countries of Europe that are experiencing various stages of financial and economic development. The empirical research will aim to define the link that exists between finance and growth in the proposed countries and will be finalized by investigating the feedback relationship between the two variables. For this purpose, in accordance to the main academic studies, a wide range of indicators of finance and growth will be analysed and a variety of econometric tests will be employed.

Keywords

Financial Development, Economic Growth, causality, emerging economies

1. Introduction

Financial development is sustained by the quality of the services offered by the financial intermediaries. The higher the efficiency of the latter, the more developed a financial system is (Lin, 1981). On the other hand, economic growth is achieved when the productive capacity of a country expands over time (King and Levine, 1993). Inevitably, the better a financial system is performing, the higher the returns from investments are, thus, more capital is available for the production of goods and services. Consequently, financial development causes economic growth.

It is noticed though, that despite the vast amount of empirical work devoted to this issue, findings do not always suggest a causality relationship. Moreover, the research is mainly focused on the well-developed countries, leaving on shade the emerging economies where financial development and economic growth are in their earliest stage. To this extent, conducting a research in the developing countries is a necessity; further light will be shed on the internationally discussed "finance-growth" relationship and specific conclusions will be drawn for those economies that have not captured the attention of the literature. For reasons that will be explained later on in this paper, two emerging economies of Europe are proposed for investigation, namely, Albania and Turkey.

This paper is organized as follows: Section 2 provides a brief review of the academic literature devoted to the "finance-growth" nexus whereas Section 3 outlines the main motives of the proposed research and the proposed methodology. The paper concludes with Section 4.

2. Literature Review

The main body of literature assumes uni-directional causality from financial development to economic growth. According to Dornbusch and Reynoso (1989), financial institutions have a positive impact on economic growth since they offer a numerous investment opportunities to individuals. Through diversification, financial intermediaries allow the allocation of funds into efficient investments, thus, improve the economy's performance (Ang, 2008).

Levine (1997) highlights the two main theoretical views on the "finance-growth" relationship. The first one, which is known as the neoclassical growth theory, is focused on the capital allocation function of the financial intermediaries and explains how it affects growth (i.e. by transferring funds among different industries). The second growth theory deals with productivity, thus, it is focused on other functions of financial sector that improve productivity or introduce in the market new innovative products. However, there are other studies that do not necessarily agree with the above considerations. For instance, King and Levine (1993) highlight only innovation as the main channel through which financial development accelerates economic growth.

Other scholars though, claim that causality may run in the opposite direction, in other words, economic growth may cause financial development. Indeed, this theoretical view was firstly introduced by Patrick (1966) who argues that financial development follows economic growth as a result of an increased demand for financial services.

According to Dushku (2010), the higher the growth rate of the real national income, the greater will be the demand by enterprises for external funds, more precisely, financial intermediation. Therefore, the financial system expands to fully supply economic growth.

Empirical models that examine the relationship between financial development and economic growth evolve from earlier formal models (i.e. neoclassical growth models) that simply establish if there is a relationship between finance and growth, to recent ones that aim to investigate the direction of the causality between them. Moreover, depending on the focus (single country or group of countries analysed), there are different econometric techniques/approaches employed. Still, as it will be noticed in the following review of empirical literature, findings are inconsistent.

Demetriades and Law (2006) employed a formal neoclassical model (Cob-Douglass production function) to examine the importance of the financial sector's quality in the economic growth of 72 countries. The authors concluded that the latter is affected by both the quantity and the quality of the financial sector's services. King and Levine (1993) performed a cross-section regression in 80 countries, in order to define the finance-growth relationship. They introduced 2 new indicators of growth (despite the real GDP growth): the rate of capital accumulation and improvements in the economic growth. In contrast, Rioja and Valev (2004) failed to find such relationship when various low- and middle-income countries were analysed. Their findings suggest that only in high-income countries financial development significantly affects economic growth.

Levine and Zervos (1998) studied the effects of both banks and financial markets in economic growth, thus, introduced additional indicators of financial development. Their findings suggested that banks and markets are complements in affecting positively economic growth. Rosseau and Wachtel (1998) concluded in a long- run relationship between finance and growth when Augmented Dicky- Fuller and Philip Perron tests were employed in 5 developed countries. Finally, the Granger causality test was used, whose findings suggested that are the intermediaries that lead to economic growth but failed to find feedback relationship between them.

Other researchers preferred combining both cross-border regression and time series analysis while investigating the "finance-growth" link. Beck and Levine (2001) employed the Generalized-Method-of Moments (GMM) technique in a panel of 40 countries and 146 observations and concluded that both stock markets and banks have a positive impact on growth. However, as criticized by Demetriades and Andrianova (2004), the average period of 5 years per country, assumed in the growth regression is not enough to estimate a long-run relationship, thus, their findings may not be accurate.

Recent academic research attempts to investigate developing countries as well. Results from the empirical work of Dushku (2010) suggest that there is a bidirectional relationship between financial development and economic growth in Albania, in the long run. In the short term though, the relationship could not be clearly established. Consistently, bi-directional causality (but both in the short and in the long term) was achieved from Demirhan et al (2011) when the same investigation was performed for Turkey.

3. Motives of the proposed research and the proposed methodology

Reviewing the international literature on the finance-growth link and observing its limited attention on the emerging economies, it becomes obvious that further attention should be paid to the latter. To this extent, this paper proposes that further investigation should be performed for two developing countries of Europe, namely, Albania and Turkey. The main motives behind the proposed research are explained as following:

- The existing empirical work for Albania and Turkey is insufficient. Moreover, it relies on a limited set of data, thus, the reliability of findings is questionable. Consequently, conducting further research on these countries is a necessity. Weaknesses pinpointed during the research process or strengths identified will help improve the quality of academic research in the country and provide useful background for future attempts.
- Currently, data on the indicators of financial development and economic growth are easily accessible for both countries in the official sites of their respective Central Banks. Moreover, quarterly observations will be used, that span from the third quarter of the year 1998 to nowadays (round 56 observations), thus, are quite enough to perform an accurate investigation.
- Albania and Turkey represent two countries, equally classified as emerging economies and yet far different in terms of their financial systems' composition and economic development stage. According to the Bank of Albania (BA) stability report of the year 2011, the Albanian financial system is sustained by the banks' activity and its economy has a recent low growth rate. In contrast, the National Bank of the Republic of Turkey (NBRT) stability report of the year 2011 highlights thatTurkey has an advanced financial system and its productive capacity is ranked among the largest in the world. To this extent, comparing the empirical findings from both the countries will provide a more complete view of the "finance-growth" relationship among various stages of development of the emerging economies.

To fulfil the above objectives, a variety of econometric techniques will be employed on certain indicators of finance and growth, which are intensively met in the relevant literature. The aim of the empirical research will be to determine the direction of causality between them. With reference to the studies by Levine and Zervos (1998), Rosseau and Wachtel (1998), Atakan et al (2010) and Dushku (2010), the main determinants of financial development are those capturing the banking sector development, such as bank credit to private sector as a share of GDP, domestic credit as a share of GDP, bank capital to asset ratio (weaknesses against financial risk), time deposits as a share of GDP, broad money (M2) as a share of GDP and others. Indeed, these variables will be used in the case of Albania, where no stock exchange exists and its financial sector belongs entirely to banks. Turkey though, has a more complex financial structure, comprised by a variety of intermediaries which are also traded in the Istanbul stock exchange (Demirhan et al, 2011). To this extent, additional variables should be considered in the case of Turkey, that indicate stock market development, specifically: market capitalisation as a share of GDP, total value of stocks traded as a share of GDP, daily volume of shares traded in the Istanbul stock

exchange as a share of GDP and others. Economic growth is commonly indicated by the real GDP growth or the GDP per capita growth. However, Disbudak (2010) argues that in order to avoid omitted-variables bias, the empirical model should be enriched with third variables that affect both finance and growth, such as inflation, budget deficit, government spending, net exports, population and investment. Including these variables into the bidirectional causality test may not only alter the direction of causality between finance and growth but also the magnitude of the estimates.

4. Conclusions

The "finance-growth" link remains an unsolved puzzle despite the huge academic work devoted to it. Empirical findings are mixed and they vary according to the period under investigation, model and econometric technique employed as well as country analysed. Moreover, the research is focused on the well-developed countries where a wide range of data is easily accessible, whereas limited attention has been paid to the less developed ones, whose financial systems and economies have only experienced growth recently. This paper argues that conducting a research for two developing countries, namely, Albania and Turkey, will add to the international literature by providing more evidence on the "finance-growth" nexus as well as will particularly improve the quantity and the quality of the relevant research in the proposed countries. Moreover, the conclusions that will be drawn for Albania and Turkey will help get a general idea on the relationship between financial development and economic growth in the emerging economies. With reference to the main body of relevant literature, a variety of econometric techniques will be employed on a wide range of variables, which will aim not only in determining the relationship between financial development and economic growth but also in investigating the direction of causality between them. It is strongly believed that the findings' suggestions will provide a useful background for future research in these countries and those similar to them.

References

Ang James B. (2008) A Survey of recent developments in the literature of finance and growth. *Journal of Economic Surveys*, 22 (3), p.536-576.

Atakan, T et al. (2010) The impact of financial sector development on the economic growth of emerging countries: An empirical study for E-7 countries. *Journal of Modern Accounting & Auditing*, 6(12), p. 50-62.

BA (2011) Financial Stability Report 2011, p.1-120.

Beck, Th and Levine, R (2001). Stock Markets, Banks and Growth: Correlation or Causality. World Bank Policy Research Working Paper No. 2670; Carlson School of Management Working Paper, p. 1-24. Demetriades, P and Andrianova, S (2004) Finance and Growth: What we know and what we need to know. In Goodhart (ed) *Money*, *Finance and Growth*, Routledge.

Demetriades, P and Law, S (2006) Finance, Institutions and Economic Development. *International Journal of Finance and Economics*, 11 (3), p.245-260.

Demirhan, E et al. (2011) The Direction of Causality between Financial Development and Economic Growth: Evidence from Turkey. *International Journal of Management*, 28 (1), p.3-19.

Disbudak, C (2010) Analyzing the Bank Credit-Economic Growth Nexus in Turkey. *European Journal of Economics, Finance & Administrative Sciences*, 23, p.34-48.

Dornbusch, R and Reynoso, A (1989) Financial Factors in Economic Development. *American Economic Development*, 79 (2), p.204-210.

Dushku, E (2010) Financial Development and Economic Growth: The Albanian Case. *Bank of Albania Discussion Paper, No.32.*

King, R.G and Levine, R (1993) Finance and Growth: Schumpeter might be right. *Quarterly Journal of Economics*, 108 (3), p.717-738.

Levine, R (1997) Financial Development and Economic Growth: Views and Agenda. *Journal of Economic Literature*, 35(2), p.688-726.

Levine, R (2004) Finance and Growth: Theory and Evidence. NBER Working Paper Brown No. W 10766 .University - Department of Economics; National Bureau of Economic Research (NBER).

Levine, R and Zervos, S (1998) Stock Markets, Banks and Economic Growth. *American Economic Review*, 88(3), p.537-558.

Lin, V (1981) The Impact of the Financial Sector on Economic Development: Some cross-section evidence. *Studies in Comparative International Development*, 16(1), p.44-55.

NBRT (2011) Financial Stability Report 2011, p.1-146.

Patrick, H (1966) Financial Development and Economic Growth in Underdeveloped Countries. *Economic Development and Cultural Change*, 14(2), p.174-190.

Rioja, F and Valev, N (2004) Finance and the Sources of Growth at Various Stages of Economic Development. *Economic Enquiry*, 42(1), p.127-140.

Rosseau, P and Wachtel, P (1998) Financial Intermediation and Economic Performance: Historical Evidence from Five Industrialized Countries. *Journal of Money, Credit and Banking*, 30 (4), p.657-678.

Benchmarking Total Factor Productivity in the Regulation of Natural Gas Distribution

Ondrej Machek1

¹ University of Economics, Prague, Department of Business Economics, Faculty of Business Administration, E-mail: ondrej.machek@vse.cz

Abstract. Incentive regulation and benchmarking are becoming common tools in the tariff regulation of public utilities, including natural gas distribution companies. We present a review of existing approaches to economic regulation using benchmarking models, their advantages and disadvantages and discuss practical experience with these methods. The work also includes a comparison of the total factor productivity approach with frontier-based methods such as data envelopment analysis or corrected ordinary least squares and discussion of the differences, possible weaknesses and misinterpretation of results. Consequently, we analyze and compare the productivity development of Czech natural gas distribution companies with other industries (water and sewerage, electricity distribution and transmission. railway transport) and other Central European (Slovak, Hungarian and Polish) natural gas distribution companies.

Keywords: total factor productivity, benchmarking, tariff regulation

1 Introduction

Public utilities, which traditionally include electricity and gas distribution and transportation, water and sewerage industries, transport and communications, have some characteristics which result in government regulation of these industries. Regulation usually takes place in the case when society takes the view that competition does not bring satisfactory benefits. The main purpose of economic regulation is to achieve competitive results in an environment where competition is not feasible. In a competitive environment, the inability to keep costs below the firm's competitors will make the firm go out of business. Regulatory agencies, willing to

achieve maximum efficiency, therefore face the challenge to simulate competitive pressures, which should induce regulated companies to behave as if they were exposed to real competition.

However, traditional methods of regulation do not provide sufficient incentives for utilities to increase productivity in a quest for lower costs. For this reason, regulatory agencies employ the **regulatory benchmarking.** This approach is based on comparing a firm's performance against an ideal performance which is obtained with respect to other, similar firms. The regulated firms with better performance growth than other firms operating in the industry are rewarded by greater allowed revenues.

Regulatory benchmarking methods can be based on efficiency or productivity measurement. In economics, efficiency is the ratio of real to ideal performance, whereas productivity is the ratio of outputs to inputs.

The doctoral thesis will have rather an inductive and interpretivist character, because according to the existing literature, there still doesn't exist any academic or practical consensus about which methods perform best, nor can any benchmarking model be classified as superior to any other. We will present a review of existing approaches to economic regulation using benchmarking models, their advantages and disadvantages and discuss practical experience with these methods. The thesis will also include a comparison of the total factor productivity approach with frontier-based methods such as data envelopment analysis and discussion of the differences, possible weaknesses and misinterpretation of results. We will analyze and compare the productivity development of natural gas distribution with other industries (water and sewerage, electricity distribution and transmission, railway transport) and other Central European natural gas distribution companies.

2 Literature Review

The negative phenomena of unregulated competition in gas distribution industries (natural monopoly, buyer's ignorance, destructive competition, cream-skimming and others) are described largely by Kahn (1988), Bonbright (1961) or Phillips (1993), as well as possible negative impacts of regulation (deadweight loss to society).

Competitive pressures are simulated using performance-based regulation, whose practical aspects are well described by Lesser and Giacchino (2009). Performance-based regulation is based on a price cap or a revenue cap. For each year, the regulatory agency limits the price level or revenues. The regulated company has to reduce its costs in order to increase profits. The revenues are capped according to the inflation rate (RPI-factor) and efficiency factor (X-factor), that's why this method is also called RPI-X regulation (Beesley and Littlechild, 1989). A range of benchmarking methods exists in the literature (e.g. Jamasb and Pollitt, 2001 or Farsi et al., 2007). These methods can be divided into two main streams: benchmarking efficiency (data envelopment analysis - DEA, stochastic frontier analysis- SFA, COLS and more) and productivity (TFP approach).

We will deal mainly with the TFP branch of methods. Productivity is a basic and intuitive measure of performance. We can speak of a firm's productivity, but we can also measure the productivity of workers, machines, multiple firms, industrial sectors, national economies or even global economy. Total factor productivity (TFP) is a method of measuring productivity and its growth and it is described in detail by Fried et al (2008). Among measures based on distance function, we can cite Malmouist productivity index (Caves et al., 1982), Hicks-Moorsteen productivity index (Diewert, 1992, p. 240) and Luenberger productivity indicator (Chambers, 1996). These measures require optimization problem solving (data envelopment analysis) or regression methods and are rather theoretical, because they measure the distance from a real, but unknown frontier. Other TFP measures are based on price aggregation, such as Törnqvist productivity index (Törnqvist, 1936), Fisher productivity index (Fisher, 1922) or Bennet-Bowley productivity indicator (Bennet, 1920). These measures require data about input and output prices, but can be derived directly from empirical data. The main advantage of using TFP price-based indexes is the fact that the analysis can be carried out with a low number of observations. This seems to be the major advantage in countries with only a short history of economic regulation. TFP methods are robust and transparent. However, these indexes require price information and unlike frontier-based indexes, they cannot be decomposed into components. More detailed requirements on data when using TFP, DEA or SFA, as well as advantages and drawbacks, are described by Coelli et al. (2003, p. 22-23).

An international survey of currently used benchmarking has been made by Aiofe and Politt (2009) and benchmarking methods being used in Central Europe have been described by Machek (2011). According to these articles, benchmarking is not used in all Central European countries, and if they are, frontier methods such as DEA or COLS prevail. Experiences with using total factor productivity were described by Lawrence (2009a) or Coelli and Lawrence (2006). The literature suggests that there are only few examples of pure application of TFP approach in the world. However, many past studies have been elaborated on this subject, both by scientists and regulatory bodies.

3 Proposed Methodology

The theoretical part of the thesis will deal with the existing efficiency and productivity benchmarking models and possibilities of their practical use at the national and international level. We will discuss the advantages and disadvantages of the methods and their constraints and make a review of past international experience with their use in practice. Consequently, we will focus on the total factor productivity approach and its limitations and weaknesses. Main points of interest will be: the fundamentals of TFP approach in competitive and regulated business, the differences and their causes and possible weaknesses and misinterpretation of results; the issues of relative development of efficiency and productivity in time and different initial conditions (market concentration, government regulations, different social and economic environment etc.), and data issues in TFP, i.e. problems concerning the definition of the set of comparable firms, definition and measurement of the inputs and outputs (cost-based indexes), determination of their costs or of the reference technology (distance-based indexes) and specification of the time period. We will make a review of literature focused on the newest trends in TFP.

In the practical part, we will employ well-known efficiency and productivity benchmarking models to assess the development of Czech natural gas distribution companies and determine the X-factor and allowed revenues based on these measures. We will use MATLAB and SPSS to perform numeric calculations and statistical analyses. The productivity of distribution firms will be compared with the productivity in other regulated industrial sectors – electricity transmission, water and sewerage and transport – and with comparable firms in other Central European countries – Slovakia, Hungary and Poland.

The Czech gas transportation network is operated by Net4Gas whereas the regional distribution is ensured by 6 operators: PP Distribuce, E.ON Distribuce, RWE Energie, VČP Net, JMP Net and SMP Net. Except for two firms, all operators are members of the RWE Transgas group. The calculations will be based on corrected annual time series from 2000 to 2011. These data are available in the annual reports of surveyed companies. Concerning other industries, we will evaluate ČEPS (electricity transmission), ČEZ Distribuce (electricity distribution), Severočeská vodárenská (water and sewerage) and České dráhy (railway transport).

Incentive regulation based on TFP requires determination of input and output variables. We will consider the following output variables: the volume of distributed natural gas (in GWH) and the number of customers. The weights will be calculated using econometrically estimated multi-output Leontief cost function following Lawrence (2009).

We will consider the following input variables: operation expenses (OPEX) and high, medium and low pressure grid length (in km) and the number of transfer stations. The weights will be calculated according to the shares of inputs in total costs.

4 Conclusion

The outcomes of the thesis should be useful both for theory of regulation and for the regulatory practice. In the field of theory of regulation, the results should bring new findings about the regulatory practice in Central Europe and possibilities of further development. A detailed analysis of existing approaches and their advantages and disadvantages and limitation could be useful for the regulatory agencies, deciding what regulatory methods to adopt, and empirical results could help the regulated firms with evaluating their own productivity, identifying weaknesses and threats and negotiating with regulatory agencies.

References

Haney, A., Politt, M. (2009) Efficiency Analysis of Energy Networks : An International Survey of Regulators. *Energy Policy*. vol. 37(12), pp. 5814-5830.

Beesley, M. and Littlechild, S. (1989) The regulation of privatized monpolies in the United Kingdom. *Rand Journal of Economics*, 20(3), pp. 454–472.

Bonbright, J. (1961) *Principles of Public Utility Rates*, New York: Columbia University Press.

Caves, D.W., Christensen, L.R., Diewert, W. E. (1982) Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers. *Economic Journal*. Vol. 92(365), pp. 73-86.

Chambers, R. G. (1996) A New Look at Exact Input, Output, Productivity and Technical Change Measurement, University of Maryland, College Park.

Coelli, T., Estache, A., Perelman, S. Trujillo, L. (2003) *A primer on efficiency measurement for utilities and transport*. World Bank Publications.

Coelli, T., Lawrence, D. (2006) *Performance Measurement and Regulation of Network Utilities*. Northampton: Edward Elgar.

Diewert, W.E. (1992) Fisher ideal output, input, and productivity indexes revisited, *Journal of Productivity Analysis*, Vol. 3, pp. 211-248.

Farsi, M., Fetz, A., Filippini, M. (2007) Benchmarking and regulation in the electricity distribution sector. *CEPE Working Paper No.* 54, Zurich: ETH Zurich.

Jamasb, T. and Pollitt, M. G. (2001) Benchmarking and regulation: international electricity experience. *Utilities Policy*, 9(3):107–130.

Kahn, A. (1988) *The Economics of Regulation: Principles and Institutions*. Cambridge: MIT Press.

Fried, H.O., Lovell C.A., Schmidt, S.S. (2008) *The Measurement of Productive Efficiency and Productivity Growth*, Oxford: Oxford University Press.

Lawrence, D. (2009) JGN Gas Distribution Productivity Performance. *Economic insights*.

Lawrence, D. (2009a) Regulation of Electricity Lines Businesses, Analysis of Lines Business Performance –1996–2003. Hawker: Meyrick and Associates.

Lesser, A., Giacchino, L. (2009) Fundamentals of Energy Regulation, Vienna: Public Utility Reports.

Machek, O. (2011) Regulatory Benchmarking in Central Europe: Actual Practice and Possibilities of Development for the Energy Sector. *The Annals of the University of Oradea. Economic sciences*. 20 (1), pp. 80–86.

Phillips, C. (1993) *The Regulation of Public Utilities: Theory and Practice*, Vienna: Public Utility Reports.

Lean Six Sigma in Food Industry in time of Globalization

Ivana Nikolic

Doctoral Studies, Megatrend University of Belgrade, Serbia, Faculty of Business Studies, E-mail: ivankinikolic@gmail.com

Abstract. The purpose of this study is to explore the possibilities of applying Lean Six Sigma (L6 σ) and quality management in food industry, making better products and defining the criteria of justification for further development and investment, optimizing processes that create competitive advantage, diversity from competitors by planning new strategy, making decisions that minimize costs and explore the critical success factors that affect $L6\sigma$. The study will try to adopt qualitative and quantitate approach to explore implementation of $L6\sigma$ in food industry. The focus will be on companies in European Union, who works on implementation of $L6\sigma$ or have already implemented. Analyzing data will help us to give justifications why food industries must apply $L6\sigma$, and also give opportunity to compare countries. This will give a chance to explore the importance of usage and improvement new methodologies such as $L6\sigma$ in food industry by its /company/ positioning and its credibility to get higher level on worldwide market. Also to expand what are the benefits from $L6\sigma$ in food industry, what companies may change and to find, what are critical success factors that influence on implementation of $L6\sigma$. The main limitation of the study is the fact that $L6\sigma$ is a new quality program and the implementation in food industry is very low. Food companies can utilize the best of this quality operation to improve their business and maximize added value. With the new methodology and better process, the quality of the food will be better.

Keywords: Lean Six Sigma, Innovation/Implementation, Critical Success Factors, Food industry

1 Introduction

"Create, improve contentiously, bring innovations and be original or die". This means that in the dynamic world of global competition, where knowledge is most important, companies have to follow and adopt new technology, tools and trends on market and continuously innovate and create new products and service. With this approach they are able to survive on the market and lead the battle and win competition (Vince, 2008; Jenica *et al*, 2010).

In today's big marketplace, increased globalization and constant technological advances, innovation becomes a key success factor for many companies (Sunhilde & Simona, 2007). To survive in this world of globalization, food industry has to apply new ideas and process and pay attention to agility while try to continuing to focus on quality, cost and efficiency (Edward *et al*, 2010).

In this paper, the subject of study includes of the applicability of modern tools – Lean Six Sigma in the food industry in time of globalization. L6 σ is a business improvement methodology that maximizes shareholder value by achieving the faster rate of improvement in customer satisfaction and behavior, cost quality, process speed and invest capital (Sunhilde & Simona, 2007; Hill *et al*, 2011). New tools include changes and opportunity that goes with new approach. Using and implementing L6 σ methodology, food companies can improve business environment and there performance. It can be management approach, focus on quality and continuous improvement which is based on participation of all employees to get success on long term (Lubowe & Blitz, 2008). The tool has been successful applied in manufacturing industry – production and service and get good results in both, increasing efficiency of procedures and improving product quality (Bowen & Youngdahl, 1998; Engelund *et al*, 2009).

2 Literature Review

Thousands of years ago, ancient people started to notice that soured or contaminated food made people sick. Throughout history, various methods were implemented to preserve food and to reduce the threat of foodborne illnesses. To ensure biological survival and prevent further degradation of the environment, man came to the conclusion that it is necessary to produce only quality, clean and safe food (Kostic-Nikolic, 2005). In the modern dynamic world, the food industry was successfully developed from raw material production, procurement and handling, to manufacturing, distribution, and consumption of food products (Kovach & Cho, 2011; IFS News, 2011). Almost every person in planet relies on the national and international food supply system (Roberts, 2001). However, these developments increased the risk of food borne illnesses. A simple mistake in step of food production- manufacturing may cause a large group of people to get sick in different locations at the same time, which makes food safety one of the hottest topics in the 21 Century. (Kovach & Cho, 2011)

A new style of life, in period of globalization, provides the consumer's rights on safety, biologically valuable food, and the right on full safety with main purpose on improving of life quality/ health, beside producing new values for society and industry.

Everybody knows that the customer is always right, especially when it comes to food (Iwaarden *et al*, 2008; Kovach & Cho, 2011). At home and abroad they want

to use safe and healthy food. Usage of food additives and ingredients is more complex and responsive to technological, economic, environmental, health and normative aspects. With this appears the problem/question of whether consumption of cured food takes a risk to the health of consumers. Problem could be resolve looking at the current process and products used in procedure (Kovach & Cho, 2011).

The food industry is huge, collective, diverse business with complex network –from farming to production to packing to distribution (Kovach & Cho, 2011). Because of that, food companies couldn't be focused only on speed or on quality (Ferguson, 2007; Lane, 2008).

Every company/organization, no matter which sector they are, work on that to rebound economy. Most of them try to find ways for optimizing processes to create competitive advantage diversify by planning new strategies, making new decisions and minimize costs (Burgess, 2010). This is the reason why large industries, like food industry, try to improve and implement new tools and systems to improve quality of products and services (Psychogios & Tsironis, 2012). Implementing the techniques of Lean Six Sigma can improve efficiency and quality in many area of company and remain competitive (Hostetler, 2010).

2.1 What is Lean Six Sigma?

L6 σ is based only on methodology, but it isn't standardized model, and according to Goh (2010) it is applied in almost every sector. According to the literature (Thomson *et al.*, 2005; Stivenson, 2009) the methodology on which based is DMAIC model – include five most important issues: Define, Measure, Analyze, Improve and Control (Dobriansky, 2009; Chen & Lyu, 2009). 'DMAIC Model' is developed by Motorola in the 1980s, to achieve the highest level in the Six Sigma (Dahlgaard & Dahlgaard-Park, 2006). These five phases are linked logically and because of that this approach is structured, disciplined which is rigorous, data driven, result–oriented approach to process of improvement (Vivekannathamoorthy & Sankar, forthcoming). This methodology creates and improves process but also implements new structure and strategy (Burgess, 2010). Table 1 present the important tools used in each step of 'DMAIC Model'.

Strategic Steps	Deliverables	Tools used
Define	Project Charter or Statement of Work(SoW)	Gantt Chart/Time Line Flow Chart/Process Map Quality Function Deployment (QFD)
Measure	Base Line figures	SIPOC (Suppliers, Inputs, Process, Outputs, and Customers) or IPO (Input- Process-Output) diagram
Analyze	Identified Root Causes	Cause-and-Effect Diagram 5-Why Scatter Diagram Regression ANOVA
Improve	Selected root causes and counter measures Improvement Implementation Plan	Affinity Diagram Hypothesis Testing DoE Failure Mode Effect Analysis (FMEA)
Control	Control Plan Charts & Monitor Standard Operating Procedures (SOP) Corrective Actions	Control Charts Poka-Yokes Standardization Documentation Final Report Presentation

Table	1 – DMAIC	Methodolog y
-------	-----------	--------------

With L6 σ techniques, companies are able to identify and categories the waste, known as – "DOWNTIME + A" (Hostetler, 2010). Identifying and quantifying the waste is one of most important steps that company can implement as a solution that could improve the effectiveness of the overall process.

From this it is shown that putting these two concepts together, Lean and Six Sigma allow to remove the waste, streamline processes, and standardization of those processes to allow simultaneously decreased cost increase in quality and achieving the fastest rate of improvement in customer satisfaction (Hostetler, 2010; Polk, 2011; Crough, 2012).

2.2 The Applicability of Lean Six Sigma in the Food Industry

Successful stories of $L6\sigma$ implementation are common in general industry. Unfortunately, there has been little implementation of $L6\sigma$ in food industry (Stivenson & Jain, 2005). Sometimes applying new tools is not only the technical nature concerns, but more or less depends of company policies, procedures, culture, fierce competition, high variable material costs, regulatory requirements, quality management and etc. (Cutler, 2007). The law and regulations are most important for food industry. With license and regulations, companies are responsible to deliver safety food. For example, the European Union main food safety law is CE 178/2002 and it is applicable from the 1th January 2005 (Europa-Food Safety, 2011). And a company in this region has to implement this with local law. Compared to the automotive industry, trust and contract complexity between buyer-supplier is significantly different (Moore, 2007). The concept of achieving quality control and continuous improvement is extraordinary critical in the food industry (Nachiappan et al., 2009). The generic quality criteria classification in food industry is present in Table 2. A quality criteria is very important because it represents what are customers expecting and from the other hand what public authorities control. The customers are looking for variety of food, pleasant tastes easy to prepare (comfort criteria) but most of them look for food that will not damage the health (safety criteria) (Kovach & Cho, 2011).

Quality criteria classification in food industry		
<u>Criteria</u>	Description	
Safety/ Reliability	Food for human consumption has to be safe for the health.	
Compliance	Product and food processing must comply to the national food laws	
Product mensuration	Food industries have to deliver food packaging respecting the announced measures such as weight, size, trace element composition, fat content	
Product nutritive quality	Processed food must be delivered with its original vitamins, proteins, trace elements and all other components for a balanced diet.	

Table	2: Quality	criteria c	lassification	in food	industry
-------	------------	------------	---------------	---------	----------

Appetizing product	Smell, colors and texture have to be attractive and compliant to the local usage.
Ease of use and packaging	Ease of use is going from cooking information, prepared food to manageable packages.
Product information	All legal and practical information on the product must be available (on the package, website)
Processed in respect of SRE	All precaution for environment protection and waste management are taken during processing and for the raw material.

Due to $L6\sigma$'s importance, all industries are leaving the traditional way of operating and start to apply these new tools (Burgess, 2010; Kamran, & Sajid, 2010). But some sectors from food industries try to keep the traditional way of producing in hope that they will be different from competitors. Figure 1 shows the major $L6\sigma$ elements in food industry

Figure 1- Major Six Sigma elements in the food industry



The main task of the food industry becomes development, and implementation of scientific achievements goes when elements of customer satisfaction, expected quality attributes, process control, equipment capability and commodity. If some of these elements miss, the product will suffer and that will have big effect on customer satisfaction, complains, vague quality expectations, missing or wrong process check, high levels of waste or downtime in production, declining profitability.

Some research explored that approximately 30% to 50% of the costs in food industry are due to costs related to the slow processing speed or performance to meet customer needs and demand.

Process cycle efficiency (PCE) was improved from 5.02% to 17.46% due to the implementation of Lean practices. The data demonstrates that $L6\sigma$ tools will contribute to the implementation of the food safety system to minimize risk, improve productivity and quality of products, and reduce unnecessary waste and time.

Also using $L6\sigma$ tools provide lowest costs; process can be defined, measured, analyzed, improved, controlled and finally benchmarked through other facilities. It creates the speed of the process (reducing cycle time) and efficiency (minimum time and cost of invested capital). It reduces defects and variations in the process (Collier, & Evans, 2007). So when the defects are reducing by 10%, the speed of process will increase for 40%.

In other words, design of $L6\sigma$ is a technique that is specially focuses on supporting design and delivery of new technology. It would reduce variations, increase profit and eliminate defects and waste will consider form, fit and functions. Some companies from food industry (Table 3) successfully use $L6\sigma$ for several years to improve mainly the area of packing and supply chain. The main food standards as HACCP, ISO 9001, ISO 22000, ISO 14001, traceability can get benefit from this

activity (Lean Six Sigma in food processing, 2011).

Food industries			
<u>Company</u>	Quality Management Tool	<u>Sector</u>	
Kellogg's	Six Sigma	on packaging	
Britvic Soft Drink	Six Sigma	on packaging and logistics	
Mars	Six Sigma	on packaging	
Twining	Lean and Six Sigma	on conditioning and packaging	
Virgin Drink	Lean	Logistics	
Socopa	Traceability and Lean Management	meat processing	
Fleury's Michon	Six Sigma diagnostic through Consumer Help Desk	/	
Euroserum (Entremont)	Traceability and Six Sigma management	milk processing	
Hero France	Traceability and Lean Management	canning industry	
Coop de France	Lean Management	Agriculture	

Table 3 – Food companies that introduce some quality operations

Rung's fishery Quality Management and Lean Management	/
---	---

According to some research, business improvements that $L6\sigma$ will bring in food industry are:

• to focus on quality criteria expected from the market

• to benchmark its quality and efficiency level and to compare it with the European level

• to link financial and operational objectives through the quality criteria

• to use the future LSS ISO standard (ISO 13035)

Experiences of using L6 σ as an improvement method have demonstrated decrease in general costs from 25 to 40% without new investment. Knowing the type of criteria and level of quality that are expected, food companies can make a business plan that will take advantage and easily enter to new markets. The quality of business plan should be built on principles of L6 σ .

According to Kovach and Cho (2011), the four major sources of variations in food industry processes which $L6\sigma$ will reduce are:

1. Insufficient design margins (poor design machinery and operating environments)

2. Inherent variability of any manufacturing process (the 5 Ms-machine, methods, materials, man and measurement)

3. The measurement system's inherent variability, which could cause up to 25% of the problems and defections in productions

4. Variable productions provided by subcontractors and vendors

Cooperation between producers and other partners interested in opening a number of innovation issues that are still looking for answers. One of the outstanding issues raised in the study: the question of the justification for further development. Justification, which is at the heart of this work, should include not only economic but also social, ethical and environmental standards. By establishing direct ties with the consumer/customer, i.e. the creation of partnerships in the network of associates, some criteria of justification may not be acceptable to all participants: required selection criteria for each individual network members.

3 Research idea and justification

The subject of study includes determining the feasibility of applying the new tool – Lean Six Sigma in industry of food. The work is focused on management of innovation, with special emphasis on research and development activities, opportunities and changes in food industry. The point of importance to use innovations is positioning, and credibility that companies get - higher level on worldwide market.

Operations of companies that produce and service food products are directed towards satisfaction of stakeholders' needs, and especially consumers' needs. The basic survival of companies and adaptation to market is quality of food adapted to needs and expectations of the buyers (Djekic & Jovanovic, 2010). The connection between quality and competitiveness shows that the level of quality indicates the level of competitiveness. European market imposes the obligation to provide adequate standards in the area of food production. Also, the global environment demands that disposal of food products is done with the help of innovative marketing.

Quality is one of basic strengths of competitive advantage and differentiation in domestic and international markets (Ehigie & Mc Andrew, 2005; Djekic & Jovanovic, 2010; Psychogios *et al*, forthcoming). It is represented by the ability of enterprises to reduce operational costs, increase productivity and use innovative technology to meet the needs of customers whose long-term confidence is expected (Djekic & Jovanovic, 2010). Continuous improvement of quality system is the key of business success of every company, ability to survive in the food market and get cooperation in the world. Quality control goes through organizational adaptation to new conditions; innovations; new tools and trends.

The need for innovation is becoming more and more apparent. Innovation process has steps, beginning from problem analysis to idea generation, evolution, project planning, product development and testing the final product on market place (Polk, 2011). A modern approach in the period of globalization is to manage innovation process that involves multidisciplinary knowledge and research simultaneously: from ideas, inventions and commercial applications, all members are included in the value chain (from supplier to consumer/customer) (Knowles *et al*, 2005).

The starting point was the idea that the food companies try to apply this approach because of the benefits that they will get: become faster and more responsive to the customers and their demands; increased customer satisfaction; operate at lowest costs of poor quality; achieve greater flexibility (Wang & Chan, 2010). The poor processes directly affect client service and satisfaction (Hostetler, 2010).

With synergy of Six Sigma and Lean operations, appeared new powerful tool- $L6\sigma$ (Antony *et al.*, 2003; Naslud, 2008; Shan *et al*, 2008). The $L6\sigma$ is mix from three most adopted and implemented tools – Total Quality Management (Pshycogios & Priporas, 2007; Moosa & Sajid, 2010), Six Sigma (Iwaarden *et al*, 2008; Aboelmaged, 2010) and Lean operations (Anderson et al, 2006; Pshychogios *et al*, forthcoming). The Table 4 shows similarities and differences between these three tools. Using this approach, we can take advantages-speed of Lean and the quality of Six Sigma and create an integrate system with the goal to achieve more together that they could do separately (Nabhani, 2009; Burgess, 2010; Jenica *et al*, 2010). The L6 σ concept is balance between these two methods (Hostetler, 2010; Psyhogios & Tsironis, 2012). The innovations fallow turns and twists and that what happened with L6 σ (Burgess, 2010).

Table 4: Similarities and differences between TQM, Six Sigma and Lean

	<u>Concepts</u>	<u>TQM</u>	<u>Six Sigma</u>	<u>Lean</u>
--	-----------------	------------	------------------	-------------

Origin	The quality evolution in Japan	The quality evolution in Japan and Motorola	The quality evolution in Japan and Toyota
Theory	Focus on customers	No defects	Remove waste
Process view	Improve and uniform processes	Reduce variations and improve proccesses	Improve flow in process
Approach	Let everybody to be comitted	Project M anagemant	Project M anagemant
Methodologies	Plan, do, study, act	DIM AC method	Understand customer value, value stream, analysis, full, perfection, pull
Tools	Analitical and statistical tools	Advance statistical and analytical tools	Analytical tools
Primary effects	Increase customers satisfation	Save money	Reduce lead time
Secondary effects	Achieve customers loyalty and improve performance	Achieves business goals and improves financial performance	Reduce inventory, incerase profitability and customer satisfaction
Criticism	No tangible imrovments, recource-demanding, unclear notion	Doesn't involve everybody, doesn't improve customer satisfation, doesn't have a sistematic view	Reduce flexibility, causes congestion in the supply chain, not applicable in all industries

4 Research methodology and data

Research methodology will try to explore implementation of $L6\sigma$ in food industry. The chosen companies are from European Union and selection is based on how they are famous in the market and cover all food production/service sectors (milk, candies, meat, fruits and etc.). This will give opportunity to see which of these sectors already apply and why $L6\sigma$; what are the benefits and what has to be changed. The number of the companies that will be tested is still unknown.

The research aims to explore whether and how companies in Europe Union use a process quality management like as $L6\sigma$ in order to improve better production/service delivery processes.

Overall questions should be: In what way will application of $L6\sigma$ affect the production/service in food companies from European Union?

From this question, arise basic research questions on this topic:

• What are the critical factors that influence the implementation of L6 σ ?

• How the application of $L6\sigma$ is going to affect the structure of the company and strategic planning (long/ short term goals; operating costs; quality)?

• How effects can be measured?

• What are the opportunities and limitations from applying this quality operation?

Considering these questions, objectives of this research appear to be the following:

 \checkmark To identify companies targets and objectives in implementing L6 σ in food production/service.

 \checkmark To develop measuring criteria.

 \checkmark To investigate how present workplace will react on implementation of L6 σ and what level of change companies should do it in future.

 \checkmark To see how many companies already apply this operation and what are the advantages/disadvantages, which they get with L6 σ .

4.1 Research methods – data collection

According to literature review, research will be based on applying appropriate qualitative (interviews) and quantitative (questionnaires) methods and procedures.

The interviews are short, no more than one hour and from them you could find more about quality, production, technology, controlling and etc. By Psychogios (2007), structured interviews are very need to investigate research.

Questionnaires will consist of three sections – general features of the research (demographical questions, considering sex, are group, education, position in the company-sectors); things about corporate requirements related to the theory of $L6\sigma$ (strategic planning, quality of production, projects) and current situation of the company (quality control, implementation of new tools, standards, techniques, programs). With questionnaires will explore the efficiency from large number of data gathering rapidly (Jenkins, 2006). Questions can be marked from 1-10. Also it will consider open questions– new ideas from companies, other operations that use or disadvantages.

4.2 Way of analyzing data

Analyzing statistical data, will give research of the market, particularly the supply and demand for quality of products, as well as supply and demand of customer and critical success factors that influence on implementation of $L6\sigma$. The results will compare with the modern requirements of quality systems and standards that are developed in EU.

4.3 Research limitations

The main limitation of these study is the fact that implementation of $L6\sigma$ in food industry is something new. The food industry is based on "basic knowledge" and it belongs to low-tech industry. Limitations could be that most of companies from this sector don't have enough knowledge about this quality approach or some of them

would like to try to implement but they doesn't have people how are trained to do that.

5 Conclusion

 $L6\sigma$ can be effective tool for improving efficiency of company's process. The three main expectations of customers are quality, taste and value. The complaints from customer, in food industry, are due poor design, improper execution and low quality. Without of combination of these three components, product/service in food industry can be realized. Based on two dimensions, technology and market, it will bring us clear understanding of implementing the $L6\sigma$ in food industry, because the concept of continuous improve in food industry is extraordinary critical.

Our goal regarding above mentioned is to provide the customer satisfaction and to see how many companies follow trends on the market through usage of innovation. One tool for better food quality is philosophy of $L6\sigma$. By this innovation food industry will response on customer/market requests and will provide higher level of quality and more economical way of acting.

The method of comparison between companies aims to determine the advantages of one over another, in order to suggest installation of the highest standard.

References

Aboelmaged, G.M. (2010) Six Sigma quality: a structured review and implications for future research. *International Journal of Quality & Reliability Management*. Vol. 27, N. 3, p. 268-317.

Andersson, R., Eriksson, H. and Torstensson, H. (2006) Similarities and differences between TQM, Six Sigma and Lean. *The TQM Magazine*. Vol. 18, N. 3, p. 282-296.

Antony, J., Escamilla, L.J. and Caine, P. (2003) Lean Sigma. *Manufacturing Engineer*. Vol. 82, N. 2, p. 40-42.

Bowe, D.E. and Youngdahl, W.E. (1998) Lean service in defense of a production line approach. *Journal of Service Industrial Management*. 9,0, p. 207-225.

Burgess, D. (2010) Lean Six Sigma: Red hot relevance. Quality. Vol. 49, N. 6, p. 42-45.

Chen, M. and Lyu, J.J. (2009) A Lean Six-Sigma approach to touch panel quality improvement. *Production Planning & Control*. Vol. 20, N. 5, p. 445-454.

Christensen, M.C. (2011) The Innovator's Dilemma: The Revolutionary Book That Will Change the Way You Do Business. Harper Business, Reprint edition.

Collier, A.D. and Evans, R.J. (2007) *Operations management : goods, service and value chains.* Division of Thomson Learning, South-Western.

Crough, C. (2012) Lean Six Sigma Value- Industry ready for it?. *ConcreteProducts*, Vol. 115, N. 3, p. 6-8.

Cutler, R. T. (2007) Food manufacturing and six sigma. *Technology Managemant for Decision – Makers*, Vol. 2, N. 0, p. 20-22.

Dahlgaard, J.J. and Dahlgaard-Park, S.M. (2006) "Lean production, six sigma quality, TQM and company culture", *The TQM Magazine*, Vol.18, N. 3, p. 263 – 281.

Djekic, S. and Jovanovic, S. (2010) Orientation to the Serbian Agricultural Competitiveness of Quality. *School of business*, p. 24-31.

Dobriansky, J. (2009) Improve Your Acquisition Process with Lean Six Sigma, Part 3. *Contract Management*, Vol. 49, N. 4, p. 60-67.

Edward, D.A., Maleyeff, A. and Venkateswaran, V. (2010) Lean Six Sigma for the twenty – first century. *Proceedings for the Northeast Region Decision Sciences Institute (NEDSI)*, p. 586-591.

Ehigie, B. & McAndrew, E. (2005) Innovations, diffusion and adoption of total quality management (TQM). *Management Decisions*, Vol. 43, N. 6, p. 925-940.

Engelund, E.H., Breum, G. and Friis, A. (2009) Optimisation of large-scale food production using Lean Manufacturing principles. *Journal of foodservice*, Vol. 20, N. 0, p. 4-14.

European Commission – Health and Consumers Food (2011).

Ferguson, D., (2007). Lean and Six Sigma: The same or different? *Management Services*, Vol. 51, N. 3, p. 12-13.

George, L. M. (2002) Lean six sigma. (1st ed.). McGraw-Hill, New York.

Goh, T.N., (2010) Six Triumphs and Six Tragedies of Six Sigma. *Quality Engineering*, Vol. 22, N. 4, p. 229-305.

Hill, A.V., Weiyong, Z. and Gilbreath, H.G. (2011) Discipline your Lean Sigma programs. *Industrial Engineer: IE*, Vol. 46, N. 6, p.48-52.

Hostetler, D. (2010) Get Results: Improve Your Accounting Firm Processes Using Lean Six Sigma. *Journal of Accountancy*, Vol. 209, N. 1, p. 38-42.

IFS News - IFS Food 6 (2011).

van Iwaarden, J., van der Wiele, T., Dale, B., Williams, R. and Bertsch, B. (2008) The Six Sigma improvement approach: a transnational comparison. *International Journal of Production Research*, Vol. 46, N. 23, p. 6739-6758.

Jenica, P.A., Mihai, D. and Sorin, A. (2010) Using lean Six Sigma as a motivational tool for processes improvement. *Annals of the University of Ordea, Economics Science Series*, Vol. 19, N. 2, p. 438-422.

Jenkins, N. (2006) A Project Management Example or "guide to make projects work".

Kamran, M. and Sajid, A. (2010) Critical analysis of Six Sigma implementation. *Total Quality Management*, Vol. 21, N. 7, p. 745-759.

Knowles, G., Whicker, L., Femat, H.J. and del Canales, C.F. (2005) A conceptual model for the application of Six Sigma methodologies to supply chain improvement. *International Journal of Logistics Research and Applications*, Vol. 8, N.1, p. 51-65.

Kovach, T. and Cho, R. (2011) Better processes make Good Eats: Food industry can benefit from lean Six Sigma principles. *Industrial Engineer*, Vol. 43, N. 1, p. 36-40.

Lane, G. (2008) Lean made your way. Industrial Engineer, Vol. 40, N. 2, p. 34-38.

Lubowe, D. and Blitz, A. (2008) Driving operational innovation using Lean Six Sigma. *Business Performance Management*, Vol. 6, N. 3, p. 10-15.

Moosa, K. and Sajid, A. (2010) Critical analysis of Six Sigma implementation. *Total Quality Management*, Vol. 21, N.7, p.745-759.

Moore, M. (2007) Getting to the core – applying lean in the food industry.

Nabhani, F. (2009) Reducing the delivery lead time in a food distribution SME through the implementation of six sigma methodology. *Journal of Manufacturing Technology Management*, Vol. 20, N.7, p. 956-957.

Nachiappan, R.M., Anatharaman, N. and Muthukumar, N. (2009) Integrated Approach to Total Productive Lean Six Sigma (TPLSS) Implementation in Manufacturing Industry. *ICFA Journal of Operations Management*, Vol. 8, N. 2, p. 14-35.

Näslund, D. (2008), "Lean, six sigma and lean sigma: fads or real process improvement methods?", *Business Process Management Journal*, Vol. 14, N. 3, p. 269 – 287.

Nikolic-Kostic, S. (2005) Criteria justification cooperative developments of innovation, case study: Development and Medicinal Extract, Aromatic plants as preservative meat products. Doctoral dissertation, University of Belgrade.

Pepper, J.P.M. and Spedding, T.A. (2010) The evolution of lean Six Sigma. *International Journal of Quality & Reliability Management*, Vol. 27, N. 2, p. 138-155.

Polk, J.D. (2011) Lean Six Sigma, Innovations and Acceleration Process Can Work Together. *Physical Executive*, Vol. 37, N. 1, p.38-42.

Pratiques Lean et Six Sigma – Lean Six Sigma in Food Processing (2011).

Psychogios, G.A. and Priporas, C.V. (2007) Understanding Total Quality Management in Context: Qualitative Research on Managers' Awareness of TQM Aspects in the Greek Service Industry. *The Qualitative Report*, Vol. 12, N.1, p. 40-66.

Psychogios, G.A. and Tsironis, L.K. (2012) Towards an integrated framework for Lean Six Sigma application: lesson from airline industry. *Total Quality Management*, Vol. 23, N. 3/4, p. 397-415.

Psychogios, G.A., Atanasovski, J. and Tsironis, K.L. (forthcoming) Lean Six Sigma in a Service Context: A Multi-Factor Application Approach in the Telecommunications Industry.

Roberts, A. C. (2001) *The food safety information handbook*. (1st ed.). CT: Animprint of Greenwood Publishing Group. Inc., Westport.

Six Sigma Online. (2009) Lean six sigma - practical applications in fast food.

Saunders, M. (2003) Research Methods for Business Students. 3rd edition. Prentice Hall, Harlow.

Shan, R., Chandrasekaran, A. and Linderman, K. (2008) In pursuit of implementation patterns: the context of Lean and Six Sigma. International Journal of Production Research, Vol. 46, N.23, p. 6679–6699.

Stevenson, W. (2009) Operations Management. McGraw-Hill/Irwin, NY.

Stevenson, L. & Jain, R. (2005) Lean manufacturing in the food industry. *Food Science and Technology*, Vol. 19, p.32-34.

Sunhilde, C. and Simona, T. (2007) Lean Six Sigma and Innovations. *Fascicle of Management and Technological Engineering*, Vol. 6, N.16, p. 2525-2530.

Thompson, A.A., Strickland III, A.J. and Gamble, E.J. (2005) *Crafting and Executing Strategy: The Quest for Competitive Advantage: Concepts and Cases*, 14/e. McGraw-Hill/Irwin, New York.

Vince, G. (2008) Lean Six. Management Services, Vol. 52, N.1, p. 22-33.

Vivekannathamoorthy, N. and Sankar, S. (fothcoming) *Lean Six Sigma, Six Sigma Projects and Personal Experiences*, KCG College of Technology, Chennai India

Wang, K.F. and Chen, S.K. (2010) Applying Lean Six Sigma and TRIZ methodology in banking services. *Total Quality Management & Business Excellence*, Vol. 21, N.3, p. 301-305.

Exploring the potential of a satisfied and loyal employee as effective and best marketing and/or PR agent of his/her company

Lana Milošević¹ ¹ Faculty of Economics and Management, Slobomir P University, lanamilosevic@yahoo.com

Abstract. This study explores whether and how people oriented corporate culture can create satisfied, loyal and high performed employees who will spread positive word-of-mouth and act as the best marketing and/or Public Relations (PR) agent of their company. After the literature review on the topics: HRM. Corporate Culture, Internal Marketing, PR, and Employee Satisfaction has been done, the existence of the correlation between variables: people oriented culture, satisfaction, lovalty, and word-of mouth of the employees is explored. The research consists of qualitative (interviews with management and unstructured observation) and quantitative parts (questionnaire for employees). Hypotheses are tested and the relationship (positive correlation) between people oriented culture on one side and satisfaction and loyalty on the other side is confirmed. Also, it is established that a positive correlation exists between employees' satisfaction and loyalty and employees' positive word-of-mouth about their organizations. This research confirms the value of HRM in the process of creation one working environment of satisfied, motivated and committed employees and goes one step further, it approves the connection between HRM and Marketing and HRM's importance and contribution in the field of marketing and vice versa. This study also researches if some contemporary organizations in the Balkans are aware of how their employees can be the best marketing and PR agent of their companies if appropriate corporate culture (people oriented culture) is created.

Keywords: HRM, *People oriented culture, Employee satisfaction and loyalty, Word-of-mouth, PR*

1 Introduction

Nowadays, when the significance of humanity and different human aspects is much bigger than just gaining and maximizing the profit, the importance of Human Resource Management (HRM) also increases. Its importance is growing due to the fact that success of one organization i.e. company depends on its employees, their satisfaction and their opinion of that company.

The most valuable assets in every company are its **employees**. Some companies recognize this fact, some of them do not, and that is one of the main reasons why some companies are successful in the market and some of them are not. Those companies, who identify this truth, make effort in order to create specific corporate culture that becomes their **competitive advantage** in business (Cawood, 2008). People are the most valuable resource of every company and the source of potential success and competitive advantage but if they aren't treated in the appropriate way, they can be the main destroyer of the company and its stability (Brau et al, 2008). MacMillan and Schuler (1984) state that one of the main sources for gaining competitive advantage for one company can be its human resources (HR) practice which include: HR planning, staffing (including recruitment, selection, and socialization), assessment, reward, training and development.

There are different definitions of corporate culture and Certo (2003) states that:

"Corporate culture is a set of shared values and beliefs that organization members have regarding the functioning and existence of their organization. What type of corporate culture is present in any organization can be discovered by studying that organization's special combination of status symbols, traditions, history, and physical environment "(p.391).

In this context, **people oriented culture** of one company is a specific kind of corporate culture where employees are in the center of all activities and actions (Zuber, 2000) during the creation of a strategy and different tactical options. When threatened in that manner, employees can be the main source for differentiation, creating competitive advantage and also the most powerful weapon in the marketing battle at the market. That means, if all conditions of employees' satisfaction are fulfilled, an employee can create an extra value: make positive image for that company, spread positive word-of-mouth about company and be one of the most effective marketing (PR) instruments.

In this study issues related to **commitment to employees, people oriented corporate culture and how satisfied employees can become an effective marketing (PR) instrument and tool** of their company is explored. Whether and how people oriented corporate culture can create satisfied, loyal and high performed employees who will act as the best PR agent for their company?

It will be also explored whether or not some service organizations, like SP University, the first private university in Bosnia and Herzegovina, recognized in what manner its employees can be an effective marketing tool if people oriented culture within organization is properly created.

2 Literature Review

2.1 HRM, Internal Marketing, Corporate Culture and PR

2.1.1 Corporate reputations, corporate culture, and people oriented culture

Corporation's reputation depends on behavior of that company which needs implemented system that consists of: strategy, employees, governance, culture, controls and business processes (Dowling, 2004). The formations of **corporate reputations** can be seen at the exhibit 1.

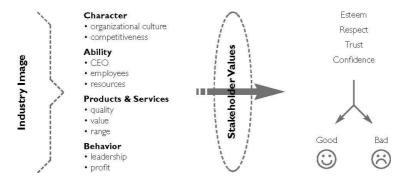


Exhibit 1 Components of Corporate Image and Identity Reputation

Source: Dowling, G. R. (2004) Corporate reputations: should you compete on yours? *California Management Review*, 46(3), p.19-36.

Dowling (2004) also states that:

"Corporate identity refers to the way that the organization presents itself to its stakeholders and answers the question 'Who are you'?" (p. 55).

Corporate identity is the **culture** of the organization but it is, according to Brock et al (2008), often disregarded by executives in the company. Cawood (2008) confirms this statement claiming that exactly culture is the thing that distinguishes successful from unsuccessful companies today; their culture contributes to their competitive advantage.

Engaged people imply enriched profits (Cawood, 2008) and they are invaluable asset of every company who can be the best marketer and creators of positive

word-of moth for their company. Condition for successful realization of this mission is to have satisfied, motivated and content employees that can be reached through good communication (via internal marketing and PR), organizational commitment to employees i.e. adequate, employee oriented culture. Cawood (2008) calls this kind of culture:

"... a **'people-profit' culture**- one where employees at every level enjoy in what they're doing, have a good relationship with management and colleagues, and are treated as the most critical contributors to their organization's success" (p.28).

2.1.2 Public relations and internal public relations

One of the most important affairs in contemporary world according to Agee et al (2005) is the growing role of management and marketing in business, attenuation of mass media, frequent communications with the individuals instead of mass audience, and attributing bigger importance to the communication with the employees. All of that has increased the role of **PR** as the fifth element of the marketing strategy besides other 4 P's: product, price, place, and promotion (Agee et al, 2005).

That's why PR becomes more specialized and PR agency focuses on only one area. Communication with employees is one of these specialized areas today (Agee et al, 2005). That changes in the world today and reduction of importance of the mass media, points that one of the aspects of PR operation can be the effective communication with the employees and using them as one useful PR instrument for outside communication with other stakeholders (consumers, community etc.).

Corporate communication includes promotion of the organization i.e. understanding it through internal and external communication (Keller and Kotler, 2006). Ferrell and Hartline (2005) in this light see PR as a part of the integral marketing communication (IMC) who prepares content that may attract the attention of their stakeholders as well as programs on the basis of which it can create and maintain good relations between them. One of these stakeholders are the employees as one of the most important part in the chain of company's business success. Companies use PR for the same reasons they use advertising: in order to convey a message to the target audience, clearly, without interference, and get the desired and proper feedback from them (Ferrell and Hartline, 2005). Ferrell and Hartline (2005) mention Starbucks as a great example of the company which creates a phenomenal competitive advantage based on adequate PR strategy and actions where they promote their fair and committed relations and attitude towards their employees. Their corporate culture is recognizable among customers and their employees don't have problems to identify with this culture. Even more, they are proud that they are a part of this company and try to share Starbucks' "messages" and outside the workplace, among family, friends and beyond.

Internal public relations (IPR) include a series of actions and activities that are meant to influence the awareness and attitudes of employees in relation to the company,

and build a positive relationship with them and two-way communication. Many of the **PR** techniques can be used for achieving these goals¹.

According to McCown (2007), an employee can be very powerful weapon that can be directed against the company itself if it does not take adequate strategic actions to exploit this resource in a way that companies can realize their goals aimed at external audiences, i.e. market. McCown (2007) stresses that the key word in the whole process of informing the internal market, i.e. employees, is the **communication** which must be twosided and carried out in order to overcome communication gaps and which satisfies common awareness.

In the sea of information that is customer oriented, the customers will prefer to choose those that come from informal sources they trust. One of these sources and very effective informal instrument of marketing communication is **viral marketing**. Semenik (2002) states that viral marketing is a very powerful appliance where new media can be used in creating a "buzz" through word-of-mouth as an effective PR tool. One example of this very effective actions has been done by T-Mobile with "flash mob" dance at Liverpool station, Trafalgar Square (13.500 people) and many other places where effects and responses of the people were amazing (Give Me My Remote, 2010) (see Figure 1. in appendix A). People talk about that in positive manner which helps in creating good company's reputations and building **corporate brand**.

2.1.3 Internal marketing and commitment to and of employees

According to Ljubojevic (2001), when it comes to attitudes of employees towards the company where they work, there are 3 the most frequently mentioned: **business satisfaction**, **business engagement** and **commitment to the organization**. **Business satisfaction** refers to the individual's general attitude towards their work. Satisfaction of internal customers is a prerequisite for the satisfaction of external customers. The key role in this has the internal market and internal marketing. **Business engagement** is the extent to which the employees get into their work, investing time and energy into it. **Commitment to the company** is the extent to which employees identify themselves within the organization and want to be an active participant in it. Then, **employees' beliefs become focused in the right realization of the mission and goals of the company and are willing to invest effort for their execution.**

¹ - Company's newspapers where all the news from life and work of the company can be announced as well as plans that company has. All this has the purpose to inform employees and realize clear and effective communication with them;

⁻ E- mail notification from PR department, and/or general and/or middle management about realized activities or those that are just ahead;

⁻ Official meeting notice in emergencies or in case of extraordinary or changed circumstances;

⁻ Employee support programs, when a company takes over an advisory function and offers advice related to everyday life of employees for example financial, psychological counseling, etc;

Internal newsletters;

Training programs;

⁻ HR programs etc (Ferrell and Hartline, 2005).

Building commitment can be realized in two ways: orientation to the individual employee and highlighting the role of the company itself. The best tool for the implementation of both is the **internal marketing** (Ljubojevic, 2001). According to Ljubojevic (2001), internal marketing appears to describe the use of marketing at the internal level, within the company. In this sense, employees are seen as internal consumers. Internal marketing is dedicated to the managing and organizing marketing activities which will create and contribute to the creation of the employee motivation and education in order to act in customer oriented manner and in accordance with corporate culture (Snell and White, 2009).

Analyzing the phenomenon of **organizational identification**, Ahearne et al (2009) point out that in addition to the basic, economic objectives of gaining the profit, there is a higher level of the whole process of exchange of goods and it's not just employee satisfaction, but also the interaction, the connection between the employee and the company. At this level the importance of organizational culture comes into force; it can be transferred to the employee through middle management in an inspiring manner.

Employees, especially those in the service sector are of crucial importance to consumers and therefore the success of the company. The largest number of consumer dissatisfaction stems from the lack of under-motivated and loyal staff work. Apart from recruitment and selection, it is needed through training programs, to orient and socialize the recruited ones in order to show those who are more successful in building commitment and productivity. Orientation is the process by which organizations help recruited ones understand the company and its culture. It is very important that employees are fully aware of the values of the company. When the newly employed people understand the culture and functioning of organizations, they can be involved in the functioning of its cultural framework. Authorization of the employees should produce two positive effects: increased satisfaction and greater commitment of employees. After all, an internal marketing points out that HR is the key factor in developing and maintaining a successful business. Efficiency of internal marketing thus becomes a prerequisite for external marketing. (Ljubojevic, 2001).

2.1.4 Internal communication, PR and marketing strategy

When we are talking about achieving an internal branding, employee satisfaction, commitment, and trust of the employee, Ferdous (2008) suggests implementation of Integrated Internal Marketing Communication Program (IIMC) in order to achieve above mentioned.

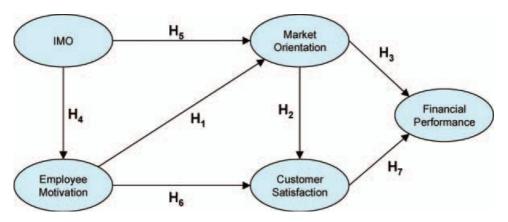
Ferdous (2008) state:

"In order to initiate the IIMC program, the initial stage consists of developing an **appropriate atmosphere for communication inside the organization** through top management support, multi-directional communication flow within the internal hierarchy and balance between the informal and formal communication... "(p.232).

Greenley and Lings (2009) in their work provide quantitative evidence that internal marketing contributes to the marketing success of the company measured with

market orientation, customer satisfaction and profit. In order to contribute to the market orientations goals, employees need to be motivated for that. Greenley and Lings (2009) confirmed their own thesis that **if motivation of employees is greater**, **the greater is the market orientation of the company**.

At the Exhibit 2, the correlation between Internal marketing orientation (IMO), employee motivation, market orientation, customer satisfaction and financial performance can be seen.



Exibit 2

Source: Greenley, G. E. and Lings, I. N. (2009) The impact of internal and external market orientations on firm performance. *Journal of Strategic Marketing*, 17(1), p.41-53.

So, according to the previously mentioned, we can state that one of the key conditions and prerequisite for the success of the company is establishing employees' satisfaction, their motivation and commitment to the company where they work.

Today, term "**modern marketing**" means integrated marketing communication (IMC) instead of promotion (Ferrell and Hartline, 2005) in order to communicate adequately with employees, community, media, business partners and other stakeholders. In this turbulent environment, businesses can survive and successfully act if they have competitive advantage which must be communicated with the target group through adequate **marketing strategy**. But good marketing strategy and brilliant marketing plan don't guarantee success in the market if execution of the marketing strategy won't properly appreciate employees of this company as equally important, the fifth element of the marketing mix (Ferrell and Hartline, 2005). This research go one step further and analyze role of employees exactly in this light, because, according to Ferrell and Hartline (2005), many companies **integrate HR and marketing department** or make them work closely particularly in employee selection and training, evaluation and compensation as

well as in employee motivation, satisfaction and commitment issues. Lee and Miller (1999) make connection between commitment to employees, strategy and competitive advantage, claiming that positioning strategies² can be achieved more successfully when company develops commitment to employees to the highest level. Many authors consider **organizational commitment to employees** (OCE) and indicate that as manifestation of OCE, appear: fair rewarding, training and development, compensation, fairness and care for employee satisfaction (Doorewaard et al, 2006; Muse et al, 2005; Lee and Miller, 2001; Lee and Miller, 1999).

After all, we can conclude that **motivated**, **capable and dedicated employees** can be created if company has develop commitment to employees and relay that people make the process (Lee and Miller, 2001). It is important to note that commitment needs to be considered as a process in two-way: without commitment to employees it isn't possible to have **organizational commitment**, i.e. to have employees who will contribute to their company.

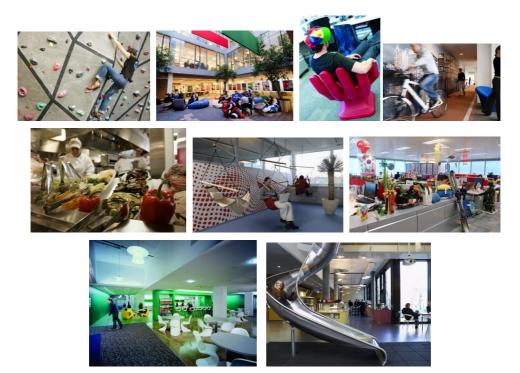
2.2 People oriented culture in function of the creation of satisfied and loyal employees and their PR activity for the company

One of the greatest examples of the implemented people oriented culture among modern companies is the example of the Google culture (The Google, 2010b). Photos below show some of the moments in everyday life at Google Company³:

 $^{^{2}}$ According to Porter (2004) those are: cost leadership, marketing differentiation and innovative differentiation.

 $^{^3}$ Here's what is written on the Google's official site about their corporate culture (The Google, 2010a):

[&]quot;Though Google has grown a lot since it opened in 1998, we still maintain a small company feel. At lunchtime, almost everyone eats in the office café, sitting at whatever table has an opening and enjoying conversations with Googlers from different teams. Our commitment to innovation depends on everyone being comfortable sharing ideas and opinions. Every employee is a hands-on contributor, and everyone wears several hats. Because we believe that each Googler is an equally important part of our success, no one hesitates to pose questions directly to Larry or Sergey in our weekly all-hands ("TGIF") meetings – or spike volleyball across the net at a corporate officer. We are aggressively inclusive in our hiring, and we favor ability over experience. We have offices around the world and dozens of languages are spoken by Google saffers, from Turkish to Telugu. The result is a team that reflects the global audience Google serves. When not at work, Googlers pursue interests from cross-country cycling to wine tasting, from flying to frisbee. As we continue to grow, we are always looking for those who share a commitment to creating search perfection and having a great time doing it."



Source: The Google (2010b) The Google Culture. *The Google Web Site* [online]. Available from: http://picasaweb.google.com/Ogogster/EverydayLifeAtGoogle?authkey=Gv1sRgCKbzi9Kw3IOx2AE&feat=flas hslideshow#5332479998701457042 [Accessed 19th November 2010].

The Google made a short official movie made as walk through Googleplex in Mountain View (CA) (The Google, 2009). In this movie it is possible to see all of the conditions and benefits enjoyed by the employees of Google^4 .

⁴ Also, in another corporate movie named "An Inside Look at Google" (The Google, 2007), one of the employees in Google claims that the difference between Google and their previous jobs is that: "You are surrounded with very bright, very enthusiastic people and things happen very quickly. The biggest frustration at some of my previous jobs was that you come up with the idea, and from the time it was hatched to the time the decisions were made; it could be years before anything was done. Here...things seem to happen in the matter of days or weeks." Here are some of the sentences employees said about Google in this movie: "Things happen", "more than just work", "startup spirit", "make life as pleasant as possible", "helpful mentoring program", "interesting job that has big impact on millions of people", "we have 20% time for working on our own projects", "we have 'tech-talks' for improving our knowledge...total free organic food in Google restaurants and fitness rooms, volleyball and basketball courts, daycare center, flexible hours for working mothers..., intimate culture enjoyed by small firms", "balance of life and work is very important at Google"... "One of the core ideas

This is a corporate movie but those employees act exactly like a great marketing and/or PR agent of their company. They are surrounded by the adequate corporate culture i.e. specific people oriented culture that they call The Google Culture and as an ultimate effect of all these culturing elements the Google employees spread positive word-of-mouth about their company. There are many videos, photos, and articles about the Google culture on the Internet that have become very popular among people i.e. Internet users as the best example of the employer in practice. People forward this to each other and spread voice about Google. There isn't better, cheaper, and more efficient marketing and PR instrument than that one⁵.

After literature review is done, it is possible to summarize what theory and some of the best practices in this field show us about this topic. Many authors stress the importance of the HRM, when creating the specific corporate culture, i.e. people oriented culture, Internal Marketing and Internal PR in the process of the creation of satisfied and loyal employees are in question. The **internal communication** is a vital factor for effective and sustainable functioning, for growth and success of the company where Internal Marketing and Internal PR have a crucial role.

On the other side, it is proved that for better functioning of the companies in today's time, the formation of a **joint Department of HR and Marketing** is even recommended. These two functions of the company must necessarily be closely linked and to effectively cooperate. Also, some of the most successful companies these days have a brilliant practice when people oriented practice is in question. During research, the great examples of the positive word-of-mouth and free PR activities that satisfied and loyal employees have done for their company are found.

Nowadays, customers respond to different **marketing incentives and efforts** less and less; in the sea of information and incentives that they are exposed to and which are similar and offer similar (or the same), they relay on **word-of mouth** from the person in which they have trust. So, one of the most important assets and sources of the competitive advantage of a company can be its employees who spread "good voice" about their company. Dowling (2004) states that our beliefs and perceptions about a company

to work here in Google are: ambitious ideas, fast responses, big achievements"... "Requirements for new candidates are: passion about what are you doing, very high energy level, commitment..."

⁵ Two years ago, CBS' journalist, John Blackstone made reportage about Google with their employees. Blackstone reports (The Google, 2008): "Free rides to work, gournet meals and massages are just a few of the benefits Google employees receive." Comments of the people, who watched the video, can be read on You Tube bellow this video and they show us that employees that were recorded in this video did the best marketing and PR job for their company. These are some of the comments: "the hell, I'll work there for free', 'I wish to be a Google worker', 'Google is going to take over the world', 'Wow. I'm inspired. Happy employees equals good work quality!', 'free food!!! OMG! Lots of other stuff to do. Free time!!! Wow! But I wonder how they manage their time?? Cool work place. SO COOL!', 'Simple awesome', 'That's what I want to live for'..."

are based on our relationship with the company, its manners in the past and what other people say about it. Employee can act in many ways as the most effective PR resource of their company and become company's worthy PR agent. The spectrum of manifestation of these activities is wide, for example, the simplest one can be when employees consume product and/or services of their firm because they are fascinated by it and recommend this to their family, friends etc. Maybe this looks like very modest contribution to the company but we can imagine what happens when the same employee goes to the competitor's house and prefer to consume/use its product/service?

3 Data and Methodology

3.1 Research hypothesis/questions

The one of the purposes of this research is to approve that there is a relationship (link) between people oriented culture on one side and the employees' satisfaction and loyalty on the other side. I.e. its intention is to prove if there is a people oriented culture at SP University (SPU) and if there is, and it is appropriately established, employees will be satisfied and loyal. Further, as a consequence of existing people oriented culture in this educational institution, satisfied and loyal employees will spread a positive word-ofmouth of the company (i.e. organization) and will act as a PR and marketing agent of it. Previously analyzed theory shows us that there is a relationship between these mentioned variables. Also, further research will tend to explore the significance and strength of these relationships.

H10 (Null): There is no relationship between people oriented culture and employee satisfaction and loyalty.

H11 (Alternative): There is a relationship between people oriented culture and employee satisfaction and loyalty.

H20 (Null): There is no relationship between satisfied and loyal employees and their actions as PR and marketing agents of their organization (employees' positive word-of-mouth about their organizations).

H21 (Alternative): There is a relationship between satisfied and loyal employees and their actions as PR and marketing agents of their organization (employees' positive word-of-mouth about their organizations).

3.1.1 Research objectives

Purpose of this study is to explore and signify advantages of implementing people oriented culture as one effective internal and external marketing tool. The objectives of this study are:

- to research the significance of people oriented culture in creating motivated, satisfied and loyal employees;
- to explore whether and how people oriented corporate culture can create the most effective PR agent of its employees;
- to determine if some organizations in service industry, like SPU, has people oriented culture;
- to examine if SPU's top management recognize benefits of people oriented corporate culture as one effective internal and external marketing tool;
- to find out if SPU management uses this tool in order to create internal as well as external additional value, and
- to identify if SPU employees acknowledge actions which are defined as people oriented and care about them.

3.2 Methodology

3.2.1 Sampling

3.2.1.1 Sampling framework

Employed workers of SPU are the target group that is the subject of interest for this study. Because of the optimization of time during testing and subsequent data processing, it was not possible to examine all the employees at the university. That's why it was chosen the corresponding sample that is considered to be representative. Beside the administrative workers and academic staff of the university were examined. Using the formula for calculating the optimal sample size it was determined that 80 questionnaires is the optimal number of surveys to obtain representative results in this study. Also, one should bear in mind that a certain deviation of the sample relative to the entire studied population may also appear.

3.2.1.2 Sampling method

Non- probability sampling, i.e. non-random sampling will be done and it will be a "snowball sampling" because members of the population are identified. As SPU has too many employees, the sample that will be chosen would be a segment that is a representative sample of the whole target population. But, some deviations of the sample in relation to whole population are possible. In the questionnaires they will be asked, mostly what, who and how in order to analyze data and properly fulfilled defined objectives of this study. All the answers that will be given in these questionnaires will be treated strictly confidentially with keeping anonymity of all these subjects without mentioning their names. Data that will be collected will be assumed as a finding about employees of the SPU as a group, not as individuals.

3.2.2 Data Collection and Instruments

Methodological approaches that will be used are both **qualitative** and **quantitative** methods of research. It will be a case study research strategy of this study since the practice of one educational institution, SPU, will be explored. Depending on answers and data that will be collected, it is planned that further statements and research findings be done. Through these questions, corporate culture, people oriented culture, satisfaction and loyalty of the employees, marketing and PR practice and especially internal marketing practice can be analyzed.

Primary and secondary research will be done as well. The primary research consists of 4 interviews and 80 questionnaires. Interviews will be done with its top managers and founders. The secondary research will be done via web site of this company as well via Internet and publications that explore this topic in theory as well modern practice in this area.

3.2.2.1 Qualitative data

Qualitative research will be done through **interviews** and through **unstructured observation**⁶. Interview consists of 16 questions (can be seen in appendix B-1). Interview includes questions (three of them) about managers perceptions of corporate culture within SPU, about satisfaction and loyalty of the employees (3 questions), and internal marketing (1 question), HRM (3 questions) and PR practice of the SP University (3 questions).

3.2.2.2 Quantitative data

Quantitative research will be done through **questionnaires**. Link with Questionnaires will be sent to the middle managers and other SPU employees via e-mails. To insure that they will answer these questions (in this e-mail) the link for the site where they can fulfill questionnaire completely anonymously will be sent to them. After interviews are completed, all these numerical data that are collected will be analyzed via SPSS. There are in total 30 questions in questionnaire which is divided into two parts (can be seen in appendix B-2). The first part consists of 5 questions of general character, i.e. **demographic questions** that help to get basic information about the profile of

⁶ Those interviews will be prepared for top managers of the SP University: the founder, general manager, rector and manager of the administration. Unstructured observation as a part of the qualitative research will be also done among other employees of the SP University in order to find out their opinion about the organization. Unstructured observation will be done especially for monitoring HRM policies and practice within SP University. Research based on the observation is possible to be done efficiently because the researcher is present in the company as an employed person for 6 years and with excluding elements of subjective assessment, it is possible to make a good and objective analysis of the current situation.

respondents. Other 25 questions⁷ in the questionnaire are related to the perception of employees regarding the issues of **corporate culture**, **employees' satisfaction**, **loyalty**, and intensity of the **positive word-of-mouth** that employees spread about company.

4. Empirical analysis

4.1 Reliability

Figure 1 Reliability Statistics

(Source: This study)

Renublity Stutistics			
	Cronbach's		
	Alpha Based on		
Cronbach's	Standardized		
Alpha	Items	N of Items	
.959	.959	25	

Reliability Statistics

As we can se at the figure 1, Cronbach's Alpha coefficient was measured in a sample of the 25 items (questionnaires) and it is extremely high: 0.959. That means that results that we will get by the means of this test will be very reliable.

Also, in the Appendix C an Item-Total Statistics is given. Values in the column Corrected Item Total Correlation on the figure 1, shows us in which degree each of this items determines (correlate) the final result. Cronbach's Alpha if Item Deleted in all cases (for all items) is near 0.95 i.e. more than 0.7 that means that there is no need to correct (remove) any of the items.

4.2 Normality

By comparing the values of the mean and the 5% Trimmed mean (Appendix C) for every of the examined variables, we can see that those values don't differ a lot. For example, the mean for total people oriented culture is 2, 43 and the 5% Trimmed mean is 2, 40. Other results can be seen in appendix C^8 . That means, according to Stojković (2001)

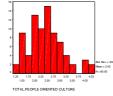
 $^{^{7}}$ Questions from 1 to 10 are about the effectiveness of the people oriented culture. Questions from 11 to 14 are about loyalty. Questions from 15 to 21 are about satisfaction. Questions from 22 to 25 are about the intensity of the positive word-of-mouth that employees spread about company and intensity of their actions as a marketing/PR agent of SPU.

⁸ Skewness (measure of asymmetric distribution) and Kurtosis (measure of the "peakedness") values for every of 4 examined variables are also given in appendix C. Values of Skewness and Kurtosis show us the nature of the distribution of the results. We can see in the appendix C that skewness is positive for all of the examined variables. That implies that the most of the results is positioned to the left of the mean and that the right end is longer (Stojkovic, 2001). As we can see from the table of descriptive statistics in appendix C, we have a positive kurtosis for variables: people oriented culture and satisfaction that means that distribution is quite peaked. Another two variables have the negative kurtosis that means that their distribution is relatively flat.

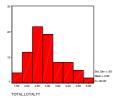
that some extreme results don't have an important influence on the mean and that there is no need for these data to be further researched.

Further, we can see the histograms below that provide a graphical view of distribution of the all of the examined variables (people oriented culture, loyalty, satisfaction and word-of-mouth). All of the histograms show us that the curve of the distribution is positively skewed for all the total variables. At the histograms, it can be seen also positive kurtosis for variables: people oriented culture and satisfaction which graphs are peaked. Relatively flat are histograms of total loyalty and total word-of-mouth. Also, in appendix C a preview of normal Q-Q plot and detrended normal Q-Q plot is given.

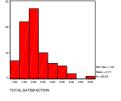
Histogram 1: People oriented culture (total)



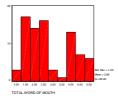
Histogram 2: Loyalty (total)



Histogram 3: Satisfaction (total)



Histogram 4: Word-of-mouth (total)



4.3 Test of hypotheses

The testing of the hypotheses will be done through comparing Pearson's correlation coefficients of the examined variables (figure 2).

	243

Figure 2: Test of correlation

	Cor	relations			
		people.oriented .culture	loyalty	satisfaction	word.of. mouth
people.oriented.culture	Pearson	1	.779**	.834**	.686**
people.onenied.eunture	Correlation	1	.11)	.034	.000
	Sig. (2-tailed)		.000	.000	.000
	Ν	80	80	80	80
loyalty	Pearson	.779**	1	.803**	.752**
	Correlation				
	Sig. (2-tailed)	.000		.000	.000
	Ν	80	80	80	80
satisfaction	Pearson Correlation	.834**	.803**	1	.738 ^{**}
	Sig. (2-tailed)	.000	.000		.000
	N	80	80	80	80
word.of.mouth	Pearson Correlation	.686**	.752**	.738**	1
	Sig. (2-tailed)	.000	.000	.000	
	Ν	80	80	80	80

Comolotiona

**. Correlation is significant at the 0.01 level (2-tailed).

(Source: This study)

Pearson's Correlation coefficient between people oriented culture and loyalty is 0.779 that means it is large positive correlation between this two variables. Pearson's Correlation coefficient between people oriented culture and satisfaction is 0.834 and it means that it is large positive correlation between this two variables. That means that hypothesis *H11 (Alternative)* will be accepted and Null hypothesis H10 should be rejected. In order to test another hypothesis, Pearson's Correlation coefficient between satisfaction and positive word-of-mouth will be observed as well as the strength of the correlation coefficient of loyalty and positive word-of-mouth. In the first case, the r= 0.738, and in another case r=0,752. As we can see, in both cases there is a large positive correlation that means that H20 (Null) will be rejected and hypothesis H21 (Alternative) will be accepted.

4.4 Collected data via questionnaires and interviews

The four hypotheses have been tested. It has been found that the relationships between people oriented culture on one side and employee satisfaction and loyalty as a consequence of the existence of people oriented culture on the other side exist. So, it is proven that there is a positive correlation between these variables, and hypothesis *H11* is

accepted. Further, researcher proved her other assumptions that if employees are satisfied and loyal, they will spread a positive word-of-mouth about their company (i.e. act as PR and marketing agents of their company). So, it is established that a positive correlation between mentioned variables and hypothesis *H21* is accepted.

Further, based on the interviews with the general management of the company, it can be seen that certain differences in their opinions exist, and also between them and employees when PR and HRM practice, the existence of the people oriented culture, and implementation of the internal and external marketing at SPU is in question. Founder's and general manager's perception about the situation is much better than the perception of the other two interviewees (manager of administration and rector) is. They all agree about the fact that it is necessary to form a HR department. All of the interviewees recognized the importance of the HR, internal marketing, and PR as well as importance of achieving the employees' satisfaction.

After all, on the basis of previously conducted research and its results, basic assumption that employees can act as an effective PR and marketing agent (spread positive word-of-mouth) if they are satisfied and loyal as consequence of the people oriented culture within organization, is supported.

4.5 Collected data via observation

SPU has financial problems nowadays and management is focused only on finding way out from these problems i.e. how to find other sources for financing and overcoming this situation. Management of SPU didn't create a strategy for managing Human Resources as well as HR department who will carry out this resource. Reason that any HR policies and practices are implemented is because of the belief that it should help SPU to become more effective. But, there is no strategic approach to this very important part within any organization (Dainty et al, 2008) and there is no **knowledge** about this. Finally, there is no persuasion that this can increase productivity and improve the organizational competitiveness and effectiveness in general. So, this level of HR policies and practices that are established at SPU cannot contribute at all the performances in general. **Recruitment and selection** politics and practices are outdated and very poor, characteristic for small, private and young firm at the Balkan. Performance objectives aren't defined, which is exactly what management by objectives by Peter Drucker strongly suggests (Abernathy, 2008). **Indirect compensation** is fully present especially when safety programs and compensation for period not worked are in question. That divides SPU as responsible employer which act in accordance with law from many other firms in region who don't respect this. **Communications** between all levels don't provide enhanced information (Abernathy, 2008).

5. Conclusions

5.1 Discussion and Implications

Today, in turbulent and unpredictable business environment where a ruthless competition rules, to be special for consumers means to have a competitive advantage. Most competitive players in the market know that the advantage can be achieved if the focus of their action is expanded. It is necessary to expand the focus from the quality of products (services), and to focus on maximizing customer satisfaction (Berman, 2005) and **on the satisfaction of their employees**. So, a place among the best will take those companies that recognize the need to look at their people as the most valuable resource in it.

If employees are treated in this way, their satisfaction and loyalty will not fail and therefore neither motivation to achieve stated goals and to overcome them (Kimpakorn and Tocquer, 2009). In such an environment, the positive effect occurs that the employees treat the company where they work as their second home and want to affirm it in the best possible way. Then, from the workers who work for the wages they become equal members of that community and spread a positive voice about it. It can even be said that they act as PR and marketing agent of their company. Analyzing the literature and observing best practices, one of the most effective ways to achieve satisfaction and loyalty of the employees is located in commitment to employees (Bae et al, 2010). That means that employees are treated within organization as the most important asset (Blanchar and Onton, 2005) and what is more important is that employees feel so. It is necessary to respect the workers, give them support and recognize their needs, to establish an adequate system of cooperation and communication in the company, adequate and timely inform them, to recognize their efforts and adequately reward them for it, to enable them to learn and be trained professionally. As a result of such a corporate culture, employees in the company feel comfortable, like their jobs, they are seen as a part of the organization, they are proud of that and want their company to do well. In such an environment employees want to promote their company to achieve even better results, spread positive news about the company or act as a PR agent when someone spreads bad stories about it.

Based on the interviews that were performed with the general management of the SPU, their beliefs about the importance of establishing adequate corporate culture are recognized. Their beliefs about the importance of association of the best HR practices, efforts and commitment to employees to achieve their satisfaction and loyalty as a prerequisite for business success, are also examined. Findings indicate that there are certain discrepancies between their perceptions of the current situation and perceptions of the employees of the University based on the answers that were given during the examination. This implies that the SPU in the future must work on more effective communication between employees and management in order to repair these problems.

After case study analysis of the variables such as people oriented culture, satisfaction, loyalty, and word-of mouth of the employees, the existence of the correlation between them is explored. Based on the questionnaires filled in by employees in both campuses, we can see that customer satisfaction and loyalty is in correlation with the corporate culture that is oriented towards the people (employees). The results of the survey show, if you achieve satisfaction and loyalty of the employees, any other positive effects, such as acting of those same workers as marketing and PR agent of the company, will not fail. In this case, therefore, a synergistic effect will be achieved and get more than it was invested. During the survey, the correlation between these variables (people oriented culture, satisfaction, loyalty and employees' word-of-mouth) has been confirmed. Both of the null hypotheses were rejected and their alternatives are accepted. But also, it was noted that the level of satisfaction and loyalty of the workers isn't at the satisfactory level. This further implies that the management of this young institution in the future has to work on improving their HR practices and increase satisfaction and loyalty of employees. In this way, management can save money that would be invested in marketing and work of PR service: if the family members do not believe in it and they are not committed to it, any other people outside will not have faith in it and talk good things about it.

5.2. Limitations and Recommendations for future research

One of the major limitations in the process of collecting secondary data is the lack of adequate literature which covers the area of Marketing and HRM analyzed in light of this topic. The theme itself is specific and needs to be fully explored so this work represents a good beginning for the further analysis of this phenomenon.

References

Abernathy, W. B. (2008) Implications and Applications of a Behavior Systems Perspective. *Journal of Organizational Behavior Management*, 28(2), p.123-138.

Agee, K. W. et al. (2005) *Public relations: strategies and tactics*. Seventh edition. Boston, Allyn and Bacon.

Ahearne, M. et al. (2009) The role of leaders in internal marketing. *Journal of Marketing*, 73(2), p.123-145.

Anon (2000) The kids are all right. *The Economist* [online], 21st December 2000. Available from: http://www.economist.com/node/455152 [Accessed 24th November 2010].

Anon (2005) Managing across cultures. *The Economist* [online], 18th November 2005. Available from: http://www.economist.com/node/5156760 [Accessed 24th November 2010].

Argenti, P. A. and Forman, J. (2005) How Corporate Communication Influences Strategy Implementation, Reputation and the Corporate Brand: An Exploratory Qualitative Study. *Corporate Reputation Review*, 8(3), p. 245-264.

Bae, J. et al. (2010) High Commitment HR Practices and Top Performers. *Management International Review (MIR)*, 50(1), p. 57-80.

Bang, H. et al. (2006) Use and Effectiveness of billboards. *Journal of Advertising*, 35 (4), p. 21-34.

Bartlett, C. A. and Ghoshal, S. (2002) Building Competitive Advantage Through People. *MIT Sloan Management Review*, 43(2), p. 34-41.

Bauer and Erdogan (2009) Organizational Behavior. New York, Flat World Knowledge.

Baumgartner, R. J. (2009) Organizational culture and leadership: Preconditions for the development of a sustainable corporation. *Sustainable Development*, 17(2), p.102-113.

Beard, P.R.J. (1992) *The MBA Experience: The Reality behind the Myth.* London, The Association of MBAS. Cited in: Clarke, K. et al. (1995) Marketing higher education: the MBA experience. *The International Journal of Educational Management*, 9(2), p. 31.

Bearden, W. et al. (2001) Measurement. *Journal of Consumer Psychology (Lawrence Erlbaum Associates)*, 10(1/2), p. 55-69.

Becker, B. E. et al (1999) Strategic human resource management at Herman Miller. *Human Resource Management*, 38(4), p.303-310.

Beltrán-Martín, I. et al. (2005) Strategic flexibility as moderator of the relationship between commitment to employees and performance in service firms. *International Journal of Human Resource Management*, 16 (11), p.2075-2093.

Berman, B. (2005) How to Delight Your Customers. *California Management Review*, 48(1), p. 129-151.

Blanchar, C. and Onton, J. (2005) Performance measures in times of change. *Employment Relations Today* (*Wiley*), 32(1), p. 25-34.

Boswell, W. R. et al. (2005) The future of HR management: Research needs and directions. *Human Resource Management*, 44(2), p. 207-216.

Bowen, D. E. and Lawler, L. L. (1992) Empowerment, why, what, how when. *Sloan Management Review*, 30(3), p. 24-29.

Brau, J. C. et al. (2008) Spirituality and Organizational Culture: Cultivating the ABCs of an Inspiring Workplace. *International Journal of Public Administration*, 31(4), p. 420-438.

Brock, M. E. et al. (2008) Corporate culture: illuminating the black hole. *People & Strategy*, 31(1), p.55-56.

Burrell, D. N. (2008) Why small private liberal Arts Colleges need to develop effective marketing cultures. *Journal of Strategic Marketing*, 16(3), p. 267-278.

Business plan. [CD-ROM]. (2008) Slobomir, Slobomir P University.

Business plan. [CD-ROM]. (2009) Slobomir, Slobomir P University.

Campbell-Allen, N. et al. (2008) Best practices in New Zealand organizations for rewarding and recognizing employee innovations and achievements. *Total Quality Management & Business Excellence*, 19(1), p.125-139.

Cawood, S. (2008) Company culture: the intangible pathway to profitability. *Employment Relations Today*, 34(4), p.27-35.

Certo, S. C. (2003) Modern management. Ninth edition. New Jersey, Prentice Hall.

Cronbach, L. J. (1957) The two disciplines of scientific psychology. *American Psychologist*, 12(1), p. 671-684.

Dainty, A. R. J. et al. (2008) Understanding employee resourcing in construction organizations. *Construction Management & Economics*, 26(11), p. 1133-1143.

Deaconu et al. (2008) Employees' satisfaction, central element of the strategy of competitive organization. *Annals of the University of Oradea, Economic Science Series*, 17(4), p.553-558.

DeNisi, A. S. and Griffin, R. W. (2005) *Human Resources Management*. Second edition. Boston, Houghton Mifflin Company.

Dewhurst, M. et al. (2010) Motivating people: getting beyond money. *McKinsey Quarterly*, 4(1), p.12-15.

Djuranović, D. and Todorović, J. (2004) *Management*. First edition. Slobomir, Slobomir P Univerzitet.

Doorewaard, J. A. C. M. et al. (2006) The effects of HRM practices and antecedents on organizational commitment among university employees. *International Journal of Human Resource Management*, 17(12), p.2035-2054.

Dowling, G. R. (2004) Corporate reputations: should you compete on yours? *California Management Review*, 46(3), p.19-36.

Dutton, J. E. et al. (2008) Giving commitment: employee support programs and the prosocial sense making process. *Academy of Management Journal*, 51(5), p. 898-918.

Edirisooriya, G. (2009) A market analysis of the latter half of the nineteenth-century American higher education sector. *History of Education*, 38(1), p. 115-132.

Fang Lee, C. et al. (2005) Outsourcing HR as a competitive strategy? A literature review and an assessment of implications. *Human Resources Management*, 44(4), p.413-432.

Ferdous, A. S. (2008) Integrated internal marketing communication (IIMC). *Marketing Review*, 8(3), p.223-235.

Ferrell, O. C. and Hartline, M. D. (2005) *Marketing Strategy*. 3rd edition. Mason, Thomson Learning.

Fox, K. F. A. and Kotler, P. (1985) *Strategic Marketing for Educational Institutions*. Englewood Clifs, Prentice Hall.

Gabrijan, V. et al. (2010) Internal and external market orientation as organizational resources- consequences for market and financial performance. *Market*, 22(2), p. 223-241.

Gilmore, A. (2003). *Services marketing and management*. First edition. London, Sage Publications.

Give Me My Remote (2010) "*Hey Jude*" *Sing Along in Trafalgar Square for T-Mobile* [online]. Newton, Give Me My Remote. Available from: http://www.givememyremote.com/remote/2009/05/04/hey-jude-sing-along-in-trafalgar-square-for-t-mobile [Accessed 19th May 2010].

Greenley, G. E. and Lings, I. N. (2009) The impact of internal and external market orientations on firm performance. *Journal of Strategic Marketing*, 17(1), p.41-53.

Hamilton, J. and Wines, W. (2009) On changing organizational cultures by injecting new ideologies: the power of stories. *Journal of Business Ethics*, 89(3), p. 433-447.

Harris, M. S. (2009) Message in a Bottle: University Advertising During Bowl Games. *Innovative Higher Education*, 33(5), p. 285-296.

Hill, C.W.L. and Jones, G.R. (1997) *Strategic management: An integrated approach*. Boston, Houghton Mifflin Company.

Hirschberg, J. et al. (2005) Descriptive Methods for Cross-Section Data. *Australian Economic Review*, 38(3), p. 333-350.

Ho, H. and Hung, C. (2008) Marketing mix formulation for higher education- an integrated analysis employing analytic hierarchy process, cluster analysis and correspondence analysis. *International Journal of Educational Management*, 22(4), p. 328-340.

Ivy, J. (2008) A new higher education marketing mix: the 7Ps for MBA marketing. *International Journal of Educational Management*, 22(4), p. 288-299.

Ivy, J. and Naudé, P. (2004) Succeeding in the MBA Marketplace: Identifying the underlying factors. *Journal of Higher Education Policy & Management*, 26(3), p. 401-417.

Jamrog, J. J. and Overholt, M. H. (2004) Building a Strategic HR Function: Continuing the Evolution. *Human Resource Planning*, 27(1), p. 51-62.

Keller, K. L. and Kotler, P. (2006) *Marketing management*. 12th edition. New Jersey, Prentice Hall.

Kimpakorn, N. and Tocquer, G. (2009) Employees'_commitment to brands in the service sector: Luxury hotel chains in Thailand. *Journal of Brand Management*, 16(8), p. 532-544.

Kotler, P. (2003) Marketing Management. 11th edition. New Jersey, Prentice Hall.

Kotler, P. et al. (2005) *Principles of Marketing*. 4th European Edition. Essex, Prentice Hall.

Lee, J. and Miller, D. (1999) People matter: commitment to employees, strategy and performance in Korean firms. *Strategic Management Journal*, 20(6), p.579-593.

Lee, J. and Miller, D. (2001) The people make the process: commitment to employees, decision making, and performance. *Journal of Management*, 27 (2), p.163-187.

Ljubojević, Č. (2001) *Management and Marketing of Services*. Second edition. Belgrade, Želnik.

Lovrić, M. et al. (2004) *The methods of Statistical analysis*. 14th edition. Belgrade, The Faculty of Economics.

MacMillan, I. C. and Schuler, R. S. (1984) Gaining competitive advantage through Human Resources Management practices. *Human Resource Management*, 23(3), p. 241-255.

McCown, N. (2007) The Role of Public Relations with internal activists. *Journal of Public Relations Research*, 19(1), p.47-68.

Meisinger, S. R. (2005) The four Cs of the HR profession: Being competent, curious, courageous and caring about people. *Human Resource Management*, 44(2), p.189-194.

Muse, L. A. et al. (2005) Commitment to employees: does it help or hinder small business performance? *Small Business Economics*, 24(2), p. 97-111.

Pallant, J. (2003) SPSS survival manual, Philadelphia, Open University Press.

Palmer, A. (2000) Principles of Marketing. New York, Oxford University Press Inc.

Porter, M. (2004) Competitive Strategy. New York, The Free Press.

Rodwell, J. J. and Teo, S. T. T. (2004) Strategic HRM in for-profit and non-profit organizations in a knowledge-intensive industry. *Public Management Review*, 6(3), p. 311-331.

Russell, M. (2005) Marketing education: A review of service quality perceptions among international students. *International Journal of Contemporary Hospitality Management*, 17(1), p. 65-77.

Salai, S. and Hegediš, I. (1994) Marketing Communications. Subotica, Subotičke novine.

Saunders, M. et al. (2000) Research Methods for Business Students. Second edition. England, Prentice Hall.

Semenik, R. J. (2002) Promotion & integrated marketing communications. Mason, Thomson Learning.

Snell, l. and White, L. (2009) An exploratory study of the application of internal marketing in professional service organizations. *Services Marketing Quarterly*, 30(3), p.195-211.

Stojković, M. (2001) Statistika. Fifth edition. Subotica, Ekonomski fakultet.

The Google (2007) An Inside Look at Google. *You tube* [online]. Available from: http://www.youtube.com/watch?v=wsnSBhWEyK4&feature=related [Accessed 21th November 2010].

The Google (2008) The 'Google' Life. *You tube* [online]. Available from: http://www.youtube.com/watch?v=_QqT38QRA84&feature=related [Accessed 21th November 2010].

The Google (2009) Life at the Googleplex. *You tube* [online]. Available from: http://www.youtube.com/watch?v=eFeLKXbnxxg&feature=related [Accessed 21th November 2010].

The Google (2010a) The Google Culture. *The Google Web Site* [online]. Available from: http://www.google.com/corporate/culture.html [Accessed 19th November 2010].

The Google (2010b) The Google Culture. *The Google Web Site* [online]. Available from: http://picasaweb.google.com/Ogogster/EverydayLifeAtGoogle?authkey=Gv1sRgCKbzi9 Kw3IOx2AE&feat=flashslideshow#5332479998701457042 [Accessed 19th November 2010].

Zuber, A. (2000) People-oriented work environment promotes a 'winning workplace'. *Nation's Restaurant News*, Monday October 16 2000, p. 1.

Understanding and improving employee engagement in organizations – An empirical study in Indonesia

Hary Febriansyah^{1,2}

1: Institute for Organization and Learning-Innsbruck University School of Management-Austria 2: People and Knowledge Management Research Group-School of Business and Management-Institute Technology Bandung-Indonesia (hary.febriansyah@uibk.ac.at / hary@sbm-itb.ac.id)

Abstract. Human resources management requires programs and strategies in order to strengthen and improve the employee's competencies, capabilities and, finally their performance. However, besides the individual formal qualifications and informal competencies, the commitment and engagement of employees for their organization and workplace plays a crucial role. This allows a major contribution for research on HRM in Indonesia and to the concept and empirical knowledge of employee engagement in a very different context from the Western world. Therefore, the main objective of this research project is develop a more sophisticated theoretical and empirical to understanding of employee engagement in the context of Indonesian organizations. According to previous empirical research on employee engagement by (Febriansyah et al., 2010), the implementation of quantitative methodology is characterized by a lack of a complexity and differentiated understanding and exploration of employee engagement. The researchers have a limited access to gather deeper insights from respondents on everyday experiences and meanings of employee engagement. Applying of qualitative methods is considered to offer complementary explorations and different knowledge on employee engagement. It provides deeper information on an issue – that is, through behaviours, beliefs, opinions, emotions, and relationships among individuals. When used along with quantitative methods, *qualitative research can help us to interpret and better understand the* complex reality of a given situation and the implications of quantitative data.

Keywords: *Human resources management, employee engagement, quantitative and qualitative method*

1 Introduction

Accepting the importance of human resources and human resources management for competitiveness and performance of organizations, we have to recognize that managing people is a particularly complex and dynamic task because the position of human resource management is at the interface between the personal systems of the employees and the structural system of a company (Koefler and Sumetzberger, 2004). Human resources management requires programs and strategies in order to strengthen and improve the employee's competencies, capabilities and, finally their performance (Schuler, 2000). However, besides the individual formal qualifications and informal competencies, the commitment and engagement of employees for their organization and workplace plays a crucial role. In fact, the aspiration to involve, engage and win commitment from employees, has been long high on the agenda of a selected portion of enlightened human resources management (Storey et al., 2008). Employee engagement is one step ahead of commitment and involvement of employees in companies (Vance, 2006). Gallup (2006 cited in Storey et.al, 2008; p. 300) has elaborated the understanding of engaged employee as 'those who work with a passion and feel a profound connection to their company and drive innovation and move the organization forward'. They work harder, are more committed, and are more likely to go 'above and beyond' the requirements and expectations of their work (Lockwood, 2007). On the other hand, these employees also tend to feel that their work actually positively affects their physical health and their psychological well-being (Crabtree, 2005). Over the years, employee engagement has become a hot and popular topic in Western countries. Meanwhile, there is a considerable amount of empirical studies on employee engagement in organizations located in the Western world (Bakker et al., 2008; Basikin, 2007; Engelbrecht, 2006; Febriansyah, 2010; Febriansyah et al., 2010; Halim, 2007; Hallberg and Schaufeli, 2006; Harter et al., 2002; Kahn, 1990; Luthans and Peterson, 2002; Maslach et al., 2001; May et al., 2004; Saks, 2006; Schaufeli et al., 2002; Schaufeli et al., 2001; Shaw, 2005; SWA and Hay-Group, 2007; Titus, 2010). However, the role of employee engagement in different national cultures and emerging markets, like Indonesia, is hardly explored (Basikin, 2007; Bhattacharya and Mukherjee, 2009; Febriansyah, 2010; Febriansyah et al., 2010; Halim, 2007; Mohapatra and Sharma, 2010; SWA and Hay-Group, 2007; Titus, 2010). Empirical studies or surveys on employee engagement in Indonesian organizations are still rather scarce. This allows a major contribution for research on HRM in Indonesia and to the concept and empirical knowledge of employee engagement in a very different context from the Western world. Therefore, the main objective of this research project is to develop a more sophisticated theoretical and empirical understanding of employee engagement in the context of Indonesian organizations. Therefore, this study will answer the following research questions: (1) How can we conceptualize employee engagement in the context of Indonesian organizations? (2) What are the essential elements or factors that influence the role and functions of employee engagement in human resources management in Indonesian organizations? (3) How can Indonesian organizations improve employee engagement?

Therefore, this study will contribute to research on HRM at the following levels: (1) It applies concepts and theories of employee engagement and reward management usually developed in and for highly developed and industrialized Western countries in

an Indonesian context. The study will also reflect on the value and accurateness of these theories in a very different economic, social, and cultural context, (2) From an empirical perspective, this research project contributes to exploring the meaning of employee engagement and the essential factors that can influence employee engagement in Indonesian organizations by collecting, processing, and interpreting qualitative and quantitative results. Particularly, it explores the influence of reward management to increase the engagement of employees at different hierarchical levels, (3) These results can form the basis for improving strategies and policies of employee engagement.

When employees are highly engaged, a number of positive outcomes will appear. Some of the outcomes of engagement link to other employee affective responses to work and team working, willingness to learn and develop self competences, related to handle the conflict and stress easily, creative ideas and problem solving, high motivation. Moreover, employee engagement has also a positive influence on employee participation in company programs, retention, and acceptance of change, confidence and proud of the organizations (organization identity). We can expect a correlation between personal outcomes a high number of engaged employees; moreover the performance of organizations will be increase. Baumruk (2006) states that companies in which sixty percent (or more) of the workforce is engaged have on average a five-years total return to shareholders (TSR) of more than twenty percent. This is comparison to companies where only forty-sixty percent of the employees are engaged; have an average TSR of about six percent. Tower-Perrin (2007) concludes that organizations with a high percentage of engaged employees increase their operating income from year to year. Harter et al., (2002) conduct a meta-analysis to examine the relationship at the business-unit level between employee satisfactionengagement and the business-unit outcomes of customer satisfaction, productivity, profit, employee turnover, and accidents.

2 Literature Review

Over the years, employee engagement has become a popular concept in human resources management (e.g. Storey et al., 2008). This is mainly due to an increasing recognition that employee engagement is an essential way towards high performance and business success of organizations (e.g. Cleland et al., 2008; Levinson, 2007). Therefore, consultants and companies, particularly HR-managers, have been trying to enable and strengthen employee engagement. (Kahn, 1990) was one of the first academics who presented a definition of engagement. He defines of engagement as 'the harnessing of organization members' selves to their work roles; in engagement, people employ and express themselves physically, cognitively, and emotionally during job performances' (p.694). Shaw (2005) defines employee engagement as 'translating employee potential into employee performance business success'. Hallberg and Schaufeli (2006) define engagement as 'being charged with energy and fully dedicated to one's work'. Lockwood (2007) defines 'employees who are engaged will work harder, more committed, and are more likely to go 'above and beyond' the requirements and expectations of their work'. All in all these definitions relate to individuals' involvement, satisfaction and enthusiasm for work. However, there are

also conflicting and ambiguous aspects and focuses of using of the term engagement in the academic literature.

According to (Macey and Schneider, 2008) definitions of engagement sometimes refer to a psychological state or even as a general personal disposition or attitude towards one's work (trait approach) whilst others are based on the behaviour it produces. In fact, with the exception of Shaw (2005), all definitions refer to employee engagement as an individual state of mind. The concepts of (Kahn, 1990; Lockwood, 2007; Maslach et al., 2001) also refer to a specific behaviour that characterizes engaged employees. That means they connect individual state and behaviour. Some definitions strongly emphasize the behavioural aspects of employee engagement: (Macey and Schneider, 2008) define engagement as 'discretionary effort or a form of in-role or extra-role effort or behaviour', involving innovative and adaptive performance and going 'beyond preserving the status quo, and instead focus on initiating or fostering change in the sense of doing something more and/or different'. These understanding make clear that engaged employees demonstrate a specific, positive behaviour at the workplace. They show a kind of extra performance and employers respect and reward these employees by providing a positive and supporting working environment consisting of attractive working conditions and encouraging working culture. When employees are highly engaged, a number of positive outcomes will appear. Some of the outcomes of engagement link to other employee affective responses to work and team working, willingness to learn and develop self competences, related to handle the conflict and stress easily, creative ideas and problem solving, high motivation.

Moreover, employee engagement has also a positive influence on employee participation in company programs, retention, and acceptance of change, confidence and proud of the organizations (organization identity). We can expect a correlation between personal outcomes a high number of engaged employees; moreover the performance of organizations will be increase. Baumruk (2006) states that companies in which sixty percent (or more) of the workforce is engaged have on average a five-years total return to shareholders (TSR) of more than twenty percent. One implication is that changes in management practices that increase employee satisfaction may increase business-unit outcomes, including profit. Employee engagement is mainly seen as a normative concept to create and maintain high job performance. It is usually characterized as an individual state or/and behaviour and related to the context of organizational structures (e.g. job design), cultures or group and team dynamics. Employee engagement can be characterized as part of modern, progressive HRM. This is connected to a strong belief of scientific and practical literature in the positive economic and social effect of employee engagement on organizations.

3 Proposed Methodology

Mixed methodology (MM) has emerged as an alternative to the dichotomy of qualitative and quantitative traditions during the past 20 years. Mixed methodology based on research has been called the third path (Gorard and Taylor, 2004), the third research paradigm (Johnson and Onwuegbuzie, 2004) or the third methodology

movement (Teddlie and Tashakkori, 2009). Probably the most commonly occurring MM combination in literature is *close-ended questionnaires and qualitative openended interviews*. This combination allows for the strengths of each strategy to be combined in a complementary manner.

3.1 Quantitative method

The instrument: Resources of instrument on this study originally from Development Dimension International (DDI), written by Bernthal (2005); Institute for Employment Studies (IES), written by Robinson (2007); and Center for Human Resources Strategy, The Rutger University, written by Castellano (2010). DDI measures employee engagement using a validated, standardized survey instrument: The E3SM: A Measure of Employee Engagement. Meanwhile, the IES develops a diagnostic tool using a regression model, to identify the most influential inputs to employee engagement. The research suggests that a sense of feeling valued by and involved with the organization is the key driver (Robinson, 2007). Updating the literature reference, utilization of Castellano's concept (2010), he delivers about a new framework and the drivers of employee engagement in the pure of academic's field. The main variables of this instrument are representing of ten indicators of human resources management components on employee engagement (Bernthal, 2005; Robinson, 2007; Castellano, 2010). Each indicator of this study contains three items; and the questionnaires ask the respondents to rate their opinion using six of Likert scale, starting from 1-6 (strongly disagree – strongly agree). In total, there are six items for respondents' profile and thirty three items for the main instrument.

3.2 Qualitative method

Following the research philosophy (interpretive, qualitative paradigm) as explained before, this study mainly relies on face-to-face, in-depth interviews. An interview is a tool of methodology in qualitative research where it is used to understand fundamental and deeper meanings in specific interactions between people (Fontana and Frey, 2005). Cohen et al., (2007) claim that face-to-face/personal interviews are the best suit for data related to experiences, opinions, values, and feelings. Although criticized by some scholars (e.g. (Silverman, 2008); (Alvesson and Skoeldberg, 2009)), the most dominant strategy for collecting qualitative data still are interviews. An interview involves one person (the interviewer) asking questions to another person (the interviewee/the informants). While interviews provide a context in which participants can ask for clarification, elaborate on ideas, and clarify perspectives in their own words, interviewers can also use questions to lead interviewee responses (Mack et al., 2005). In gualitative research, open-ended gualitative interviews are more common than closed-ended quantitative interviews. This interview form emphasizes non-directive and general questions to allow interviewees to construct their own interpretations of social phenomena (Teddlie and Tashakkori, 2009).

4 Conclusions

Modern organizations expect their employees to be proactive, show initiative, and take responsibility for their own professional development and to be committed to high quality performance standards. They need employees who feel energetic and dedicated – (i.e., who are engaged with their work) (Bakker and Leiter, 2010). Engaged employees deliver better performance, which is critical for business success. They understand their role in the business strategy, have a strong connection and commitment to the company, are more involved, and strive to go above and beyond in their jobs (Aon-Hewitt, 2012).

References

- Alvesson, M., & Skoeldberg, K. (2009). *Reflexive methodology : new vistas for qualitative research* (Second ed.). Los Angeles; London: SAGE.
- Aon-Hewitt. (2012). 2012 Trends in Global Employee Engagement: Aon-Hewitt Consulting.
- Bakker, A. B., & Leiter, M. P. (2010). Work engagement: A handbook of essential theory and research. 27 Church Road, Hove, East Sussex BN3 2FA: Psychology Press.
- Bakker, A. B., Schaufeli, W., Leiter, M., & Taris, T. (2008). Work engagement: An emerging concept in occupational health psychology. *Work & Stress*, 22(3), 187-200. doi: 10.1080/02678370802393649
- Basikin. (2007). Work engagement among secondary school English in Yogyakarta Indonesia. Paper presented at the Annual AARE Conference, Perth, Australia.
- Baumruk, R. (2006). Why managers are crucial to increasing engagement: Hewitt Associates, Melcrum Publishing.
- Bernthal, P. R. (2005). Measuring employee engagement: Development Dimensions International, Inc.
- Bhattacharya, S., & Mukherjee, P. (2009). Rewards as a key to employee engagement. *ASBM Journal of Management*, *II*(1), 160-175.
- Castellano, W. G. (2010). A new framework of employee engagement *CHRS Whitepaper*. New Jersey: Rutgers, The state university of New Jersey.
- Cleland, A., Mitchinson, W., & Townend, A. (2008). Engagement, assertiveness and business performance A new perspective: Ixia Consultancy Ltd.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Abingdon, Oxon: Routledge.
- Crabtree, S. (2005). Engagement keeps the doctor away. A happy employee is a healthy employee, according to a GMJ survey. Retrieved from http://gmj.gallup.com/content/14500/Engagement-Keeps-Doctor-Away.aspx
- Engelbrecht, S. (2006). *Motivation and burnout in human service work: The case of midwifery in Denmark national institute*. Doctor of Philosophy, Roskilde University.
- Febriansyah, H. (2010). Endorsing employee engagement through human capital approach. An empirical research. Paper presented at the Management International Conference, Ankara, Turkey.
- Febriansyah, H., Kusuma Negara, M., & Welly, J. (2010). Identification of employee engagement variables. Bandung-Indonesia: School of Business and Management, ITB.

- Fontana, A., & Frey, J. H. (2005). The interview: From neutral stance to political involvment. In Denzin & Lincoln, Y. S. (Eds.), *The Sage Handbook of Qualitative Research* (3rd ed.). Thousand Oak: Sage Publications.
- Gorard, S., & Taylor, C. (2004). *Combining methods in educational and social research*. Buckingham, UK: Open University Press.
- Halim, L. (2007). Employee engagement and business success Indonesia. Retrieved from <u>http://www.watsonwyatt.com/render.asp?catid=1&id=18448</u> website:
- Hallberg, U. E., & Schaufeli, W. B. (2006). Same same but different? *European Psychologist*, *11*(2), 119-127. doi: 10.1027/1016-9040.11.2.119
- Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis. *Journal of Applied Psychology*, 87, 268-279.
- Johnson, B. R., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14-26.
- Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal*, 33(4), 692-724.
- Koefler, A., & Sumetzberger, W. (2004). *Human resource management in a knowledge-based economy*. Paper presented at the International Conference proceeding Ljubljana.
- Levinson, E. (2007). Developing high employee engagement makes good business sense. Retrieved from www.interactionassociates.com/ideas/2007/05/developing_high_employee_e ngagement_makes_good_business_sense.php at 20 December 2010.
- Lockwood, N. R. (2007). Leveraging employee engagement for competitive advantage: HR's strategic role. In Management, S. f. H. R. (Ed.). Alexandria, VA: SHRM Research Quarterly Report.
- Luthans, F., & Peterson, S. J. (2002). Employee engagement and manager selfefficacy: Implications for managerial effectiveness and development. *Journal of Management Development*, 21, 376-387.
- Macey, W. H., & Schneider, B. (2008). The meaning of employee engagement. Industrial and Organizational Psychology,, 1, 3-30.
- Mack, N., Woodsong, C., Macqueen, K. M., Guest, G., & Namey, E. (2005). *Qualitative research methods: A data collector's field guide*. North Carolina, USA: Family Health International.
- Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. Annual Review of Psychology, 52, 397-422.
- May, D. R., Gilson, R. L., & Lynn, H. M. (2004). The psychological conditions of meaningfulness, safety, and availability and the engagement of the human spirit at work. *Journal of Occupational Psychology*, 77, 11-37.
- Mohapatra, M., & Sharma, B. R. (2010). Study of Employee Engagement and its Predictors in an Indian Public Sector Undertaking. *Global Business Review*, 11(2), 281-301. doi: 10.1177/097215091001100210
- Robinson, D. (2007). Employee engagement (IES opinion). Mantell Building, Falmer, Brighton BN1 9RF: Institut Employment Studies.
- Saks, A. M. (2006). Antecedents and consequences of employee engagement. *Journal* of Managerial Psychology, 21(7), 600-619. doi: 10.1108/02683940610690169

- Schaufeli, W. B., Salanova, M., Gonz, V., Roma, A., & Bakker, A. B. (2002). The measurement of engagement and burn out: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies*, 3, 71-92.
- Schaufeli, W. B., Taris, T., Le Blanc, P., Peeters, M., Bakker, A., & De Jonge, J. (2001). Maakt arbeid gezond? Op zoek naar de bevlogen werknemer [Does work make happy. In search of the engaged worker]. *De Psycholoog, 36*, 422-428.
- Schuler, R. R. (2000). The internationalization of human resource management. *Journal of International Management*, 6(3), 239-260.
- Shaw, K. (2005). *Employee engagement, how to build a high-performance workforce:* Melcrum Publishing Limited.
- Silverman, D. (2008). *Doing qualitative research: A practical handbook*. Thousand Oaks: London: SAGE Publication.
- Storey, J., Wright, P. M., Welbourne, & Ulrich, D. (2008). *The Routledge Companion* to Strategic Human Resource Management: Routledge.
- SWA, M., & Hay-Group. (2007). 10 perusahaan terbaik dan ternyaman 2007 (Vol. April 2007). Jakarta-Indonesia: SWA and Hay Group.
- Teddlie, C., & Tashakkori, A. (2009). Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioural sciences. Thousand Oaks, CA: Sage Publications.
- Titus, A. (2010, January 2010). 2010-A year of employee engagement part I, *The Jakarta Post.* Retrieved from <u>http://www.thejakartapost.com/news/2010/01/20/2010-a-year-employee-engagement-part-i.html</u>
- Tower-Perrin. (2007). Largest ever study of global workforce finds senior management holds trigger to unleash talent potential.
- Vance, R. J. (2006). Employee engagement and commitment: A guide to understanding, measuring and increasing engagement in your organization. Alexandria, VA: SHRM Foundation.

Public human resources management role: ways of implementing organizational social responsibility strategy

Straia (Ilieș) Camelia – Olivia 1

¹ Management –Economics Departament, Faculty of Economics, Juridical and Administrative Sciences, Petru Maior University, Tg-MureŞ, Romania cami_ilies@yahoo.com

Abstract. In the global economy, public and private organizations have the responsibility to demonstrate and promote public and social responsibility. Long-term sustainability corporate demands that all organizations rethink their strategic goals and objectives from focusing on making a private or public benefit. In the public sector, human resource managers play the responsability and the critical role - that of leading and educating public entities regarding the importance of organisational social responsibility in all activities while at the same time strategically implementing healthy human resources management practices that support the entities social responsibility goals. This paper wants to describe ways and findings for public human resources managers to promote and implement organisational social responsibility within the entity: a. define organisational social responsibility; b. establish metrics for measuring the impact of the organization s organisational social responsibility practices; c. involve civil servants discover and promote organizational social responsibility; d. communicate internally and externally. e. establish sustainable, positive and pro-active relationships with other public and private organisations. Organisational social responsibility will not solve all of society's ills, but it will go along way making the world a better place. It gives everyone a reason to smile.

Keywords: corporate social responsibility, organisational social responsibility, public human resource management, public career

1. Introduction

Most definitions of corporate social responsibility describe it as a concept that integrate social and environmental concerns of the organisation in their activities and in the interaction with their stakeholders on a voluntary basis. Being socially responsible means not only respect legal expectations, but also going beyond compliance and investing "more" into public human capital, the environment and in the relations with stakeholders.

This paper aim to find issues in which the public sector can support organisational social responsability activities, developing suggestion for policymakers.

An increasing number of European public organisation, romanian public sector also, try to promote their social corporate strategies as a response to a variety of social, environmental and economic pressures. They aim to send also a signal to the various stakeholders with whom they directly interact: public servants, employees, investors, consumers, private companies and NGOs. In doing so, public organisations are investing in their future and expect that the voluntary commitment they adopt will help to increase their profitability (COM 2001 366).

In context, public human resources structures play an instrumental role in helping public organization to achieve its goals of becoming a socially and environmentally responsible entity and promote the positive actions with impact on society and environment.

Today there is considerable guidance to organisations who wish to be "the best place to work" and for entities who seek to manage their employee relationships in a socially responsible way. But, there is a dearth of information for the public human resource manager who sees the importance of embedding their organisation's corporate social responsability values throughout the entity, who wish to assist the public servants in integrating corporate social responsability into the organisation's DNA. For this point of view, human resources public mandate must be to create, communicate, implement and develope flexible public strategies, policies, programes, cultural and behavioural changes in public field and also fulfill organization's objectives to integrate corporate social responsability in all that public organisation and the civil servants do. Civil servant's engagement is not simply the public human resources management structures duty. Public human resources management need to:

facilitate the development of processes and systems;

- civil servant's engagement is a shared responsibility, of all public structures.

Public human resources management play also a significant role helping organisational social responsability to become "*the way we do things around here*" respecting always the state law. Public human resources management structures are the key partner who can guarantee that what the organization is saying publicly aligns with how people are treated within the public organization. Public human resources manager is in the position of being able to provide the tools and instruments for the executives and public managers to introduce organisational social responsability ethic and culture as part of the strategic framework of the public entity.

In the coming years as organisational social responsability becomes part of the public agenda and the factory of responsible entities, it will become a natural effort for the public managers.

2. Future trends and drivers

Organisational social responsibility has been defined first as "a concept whereby entities decide voluntarily to contribute to a better society and a cleaner environment and, second, as a process by which companies manage their relationship with stakeholders" (European Commission, 2001).

Corporate social responsability has become a priority issue on public organisations' agendas. This has changed public organisations' "*capacity to act and impact on social and environmental issues in their relationship with companies, but has also affected the framework in which organisational social responsibility public policies are designed: governments are incorporating multi-stakeholder strategies.*" (Albareda, L., Lozano, J. M., Ysa Tamyko, 2007)

Sustainable development involves activities currently supported by projects that will support public organization in the future. In a turbulent environment, subject to the constant changes, any predictions are difficult. Therefore, it is more important that public management of a state to carry out effective work in there to get results and opportunities in the future.

In this respect, it becomes critical that public organizations identify and capitalize current opportunities, which will give rise to others in the future. The latter are the words of Peter Drucker (Drucker, 1998) consequences and effects of events already occurred: *"prepare for the future that already happened."*

Sustainable development and corporate social responsibility processes are treated together in the literature. The argument is that sustainable development "*embraces three pawns of environmental protection, social responsibility and economic development*" regardless of the type of public or private organization. (White, 2009)

In this context, discussed in relation to corporate social responsibility, sustainable development concept aims to complement the overall pattern of discussion related to understanding the concept of corporate social responsibility in public organizations.

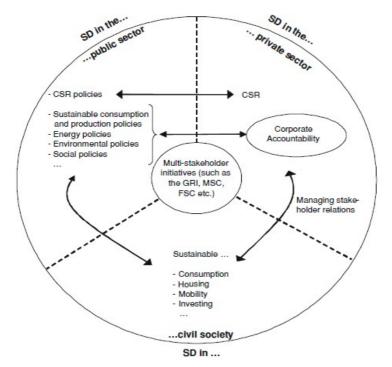


Figure 1: Sustainable development (SD) within and across the tree societal domains

Source: Steurer, R., 2009, *The role of governments in corporate social responsibility: characterising public policies on CSR in Europe*, Springer Science+Business Media, LLC. p. 54.

In our opinion, managing relationships with stakeholders is for organizations that public policies regarding corporate social responsibility is for public organizations: both are attempts to connect better corporate social responsibility and sustainable development, and often this is done in new ways of governance (corporate and public) (Fig. 1).

The roles that public organizations play in ensuring a social responsibility covers a wide range of issues related to conduct the business of corporate governance and environmental protection, issues of social inclusion, human rights and national economies.

The role of public sector in social responsibility is complex, an area in constant development. The term corporate social responsibility is not yet fully taken in the public sector entities, many of the actions were not taken explicitly into consideration like corporate social responsibility initiatives, while that actions are part of their daily concerns. There is therefore a rich experience among relevant public sector entities that remain untapped.

The following table categorizes possible government interventions regarding corporate social responsability

Mandating	'Command and control' legislation	Regulators and inspectorates	Legal and fiscal penalties and rewards	
	'Enabling' legislation	Creating incentives	Capacity building	
Facilitating	Funding support	Raising awareness	Stimulating markets	
Partnering	Combining resources	Stakeholder engagement	Dialogue	
Endorsing	Political suppor	t P	Publicity and praise	

Table 1: Public Sector Roles

Source: Nelarine C., Todres, M., Shaheena Janjuha-Jivraj (2008), Woods, A. and Wallace, J., Corporate Social Responsibility and the Social Enterprise, Journal of Business Ethics, Volume 81, Number 2, p.5

The report, Developing the Global Leader of Tomorrow, observed that "a range of human resource levers are important for developing [CSR] organizational capabilities: building these knowledge and skills through leadership development programs, career development planning, succession planning, performance management and incentive systems and competency frameworks, and seeking these knowledge and skills when recruiting new talent into the organization" (Ashridge, 2008, p. 10).

Another critical factor for the role of public human resource management in applying organisational social responsability is the orientation and implication of public servants achieving public organizational objectives. As a key driver of value in any organization, employees need to be engaged in the task of integrating organisational social responsability throughout the public entity, helping the public organisational achieve its social responsability goals and adhere to its organisational social responsability principles consistent with its strategic direction. Public or private entity's value is more and more a function of intangibles elements such as: goodwill, reputation, trust, talent, competencies and intellectual capital, etc (Report industry Canada, 2009).

Now people increasingly want to work for an organization that has a "conscience"; values build conscience. Like private organisation, public entities need to re-evaluate and re-design their organizational values in ways that are aligned with their official and public position, mission and vision. "The way things are around here", in the public formal and legal context, will be married up with public organizational targets and results. People desire to work for public entities that make a difference; sustain, afirm, comunicate, implement strong public values, infused with corporate social responsability, who generate conscience. Public sector embedding corporate social responsability throughout all it do will be the employer choice for future labour market.

These times public organisation need to be more sensitive and intentional about their corporate social responsability commitments, entities with a conscience will have to prove they are stable and viable in order to attract and develope talent – and they will have to step up their efforts to fully integrate their corporate social responsability value proposition.

Under any scenario, it is important that employees have their basic employment needs met before they are called to help the public entity achieve its organisational social responsability goals. A good workplace practices is accepted to have the critical role in building the employee value proposition and in leveraging employee organisational social responsability engagement. To have a high performing team, it is essential that people receive the proper compensation and recognition for their work.

3. Premise of integrating organizational social responsibility into public human resources management

Public human resources management peoples are highly tuned to considering corporate social responsability from both the entity "benefits" and the people "benefits" of fostering organisational social responsability alignment and integration. There are some key steps available to human resources managers who wish to develop and integrate corporate social responsability within the public organisation. They must be supported by a strong executive team and by the public managers commitment for organisational social responsability integration. As strategic partner in the public organization, public human resources managers can and should facilitate, coordinate and show their vaste expertise, but it can not dictate an organisational social responsability direction. The executive team and public managers set the tone and need to spread the talk of organisational social responsability in order to provide effective leadership. Organisational social responsability needs to be built into the public organization's strategy and into public managers performance goals, along with measurable and realistic objectives and regular performance reporting. It is a long, continous and difficult process to change in time mentalities.

Steps available to public human resources managers who desire to develop and integrate the organisational social responsability within the *"public life"* of the organisation.

3.1. Public vision, mission, values and organizational social responsability strategy development

Successful organisational social responsability strategy requires a clearly concept about public vision, mission and values. The human resource departments could initiate and support the development of public vision, mission and values foundation if one does not exist or does not explicitly address organisational social responsability. The conception of this elements needs to incorporate facts of organisational social responsability or sustainability. Where a organisational social responsability ethic has not yet implemented, the public human resource manager could exploit the need and opportunity for a vision, mission and a set of values and show how it can add outcoms to the public organization, why this could be both a good public entity strategy and a good people strategy. The public human resources manager can bring the opportunities in attention of the public manager and his executive team on what it means, and why this effort have value.

Once the public vision, mission and values are defined, the public organisation can begin the development of its corporate social responsability strategy. The role of the public human resources manager is oriented to all other steps: it is critical choice that the human resources function to be represented at *"the round table of negociation"* in the development of the organisational social responsability plan and strategic directions. Public human resources managers need to have a strong opinion about "people perspective" to contribute/support and will be highly involved in implementing key measures.

3.2. Civil service public codes of conduct

The human resource function is responsible for conceiving and implementing employee codes of conduct. Public human resources managers hold the pen in setting the principles contained in the civil servants codes of conduct. Public managers need to be also concerned about the ethical culture, looking for 100% accordance with codes of conduct which reflect the public ethical values. Codes of conduct are key tools for cultural integration of organisational social responsability prescriptions. It is important to avoid rhetoric and undefined terms such as "sustainability" and "corporate social responsability", but is very important to clearly describe the standards expected from the public servant.

3.3. Public workforce planning and recruitment

Workforce planning consists of analyzing present workforce competencies; identification of competencies needed in the future; comparison of the present workforce to future needs to identify competency gaps and surpluses; the preparation of plans for building the workforce needed in the future; an evaluation process to assure that the workforce competency model remains valid and that objectives are being met. An public organisational social responsability entity oriented means to evaluate the need for skill and competencies demanded by the economy of resource and energy scarcity, human and environmental security constraints, changing societal norms and government expectations. Public entity need to identify their main organisational social responsability competencies and gaps in the context of these changes, in this new "green economy"; the marketplace is suffering transformations that will require new green competencies and skills.

Talent management, which refers to the process of developing and integrating new civil servants in public organisation, keeping current civil servants up to date and attracte highly skilled people for the organisation, needs to consider alignment with the company's corporate social responsability vision and goals to ensure that talent is well developed according with the appropriate potential. In this area, leading corporate social responsability from private companies develop an employer brand that incorporates their organisational social responsability perspective into the employee value proposition. Entities are profiling their corporate social responsability ethics in their recruitment branding and marketing programs, promoting the benefits of working within a values-based culture. In our opinion, in public sector, civil servants volunteer programs and community involvement are the best public values who can be expressed by them.

3.4. Orientation, training and competency development in public sector

Orientation process means that public servants should have an overview of the clear line between the public organisation's vision, mission and organisational social responsability values and goals. This general orientation should be take into consideration by all levels of new public servants who need to be provided with information about organisational social responsability policies and commitments, the key organisational social responsability issues the public organisation faces and the key stakeholders with which the entity engages. How the organisations measures the organisational social responsability report, and where they can find further organisation information on corporate social responsability are elements of new civil servants orientation programs. New hires should receive a copy of the core public values and should benefit of corporate social responsability training on a regular basis.

By planning the workforce efforts, the entity will have identified the organisational social responsability key competencies the public organisation will require in future; learning plans and programs will need to address anticipated organisational social responsability competency gaps. As public management plays an important role in understanding and delivering key organisational social responsability objectives, it is vital to make this concept an important part of public management training programs.

Public human resources managers need to understand the "win-win" concept in civil servants future career and succession planning, particularly for the high performance individual. Civil servants included in efforts to advance their public career within the organisation are more motivated, more loyal and therefore, more productive employees.

3.5. Public compensation and performance management

Public human resources departments are clearly involved in setting performance standards, expectations and monitoring results to performance objectives. The most critical human resources management tool of all is the compensation and incentive program. Public human resources managers understand very well that "*you get what you pay for*". The total reward and recognition programs, including base salary, incentive pay, long term incentives and other non-monetary recognition benefits (such as award programs, employee of the month, promotions, career pathing, etc.), needs to be aligned with the entity's organisational social responsability values and strategy.

The public human resources department can help every public servant establish proper organisational social responsability targets and develop performance evaluation systems that foster organisational social responsability behaviour.

The personal objectives set by each civil servant could incorporate one organisational social responsability objective aligned with the organisational social

responsability strategy. Organisational social responsability in the public sector should be recognized in both the base job responsibilities as well as the annual performance objectives at the individual and team levels. Performance reviews could consider how the public servant has advanced their personal and the public organization's social responsability goals over the period.

3.6. Change management and public organizational culture

Public human resources managers are the *"keepers of the flame*" in building corporate culture, team building and change management processes, growing and adapting to the changing marketplace. Sometimes public organizations require outside assistance of change management professionals to help them identify an appropriate strategy when they are attempting to create significant behavioural change.

Public organisational culture shift can only be achieved and sustained if it is driven and supported effectively internally by all the civil servants. Public culture change requires also setting the tone at the top - where executives and public managers demonstrate and model the public organization's values – and then creating alignment throughout the public organization with the values you espouse to live. The values need to be reflected in all processes starting with how you attract and recruit civil servants, to decision-making, rewards and incentive programs, etc.

The public organizational culture, or "*how work gets done around here*", is a key dimension of any strong public organisational social responsability.

3.7. Civil servants involvement and participation

In every public organisation, civil servants are the key stakeholders for the development of any public strategy or program. A critical first step in public mission, vision, values and strategy development is to understand the key concerns, priorities and perspectives of all key stakeholders, particularly civil servants. It is well known that civil servants consulted and engaged in the development of new public strategies, programs and approaches sustain the good implementation of this kind of initiatiave. Often public organisations consult and engage their civil servants in the development and delivery of their community involvement and public charitable actions.

Civil servants engagement has been acknowledged as a key driver of shareholder value in a public organisation and is becoming a key for monitoring corporate performance by executives and public management.

It is expected that as organisational social responsability becomes an acknowledged component of civil servants engagement and therefore driver of public value, organisational social responsability alignment will become a critical tool for fostering public organisational success over time.

To achieve basic civil servants education and awareness, many public human resources departments become engaged in raising events and initiatives in organisational social responsability.

3.8. Organisational social responsability policy and program development

Public human resources managers are also in a position to drive policy development and program implementation in areas that directly support public organisational social responsability values. Diversity, work-life balance and flex-time policies become more and more social responsability programs directly within the public human resources manager's responsibilities. This kind of programs offer the opportunity of a platform for engaging public servants in discussions about "personal sustainability" and provide support for public servants in the areas of stress management, spirituality at work, health and fitne110ss, healthy lifestyles, etc. Another related policy could be the development of an unpaid leave program for civil servants to pursue personal projects aligned with public company values.

3.9. Public servants communication

In our oppinion, every public organisational social responsability strategy requires the development and implementation of an civil servant communication program to set the organisation direction, objectives, innovation and performance. Intranets, websites, blogs, social networking sites, videos, forums, meetings, regular team briefings, voicemails, print and electronic newsletters and other forms of social media need to be deployed to bring the public organisational social responsability message to the workforce – in ways that are attuned to the communication channels of the employee. The ultimate goal of public organisational social responsability communications should be to engage public servants in the public organisational social responsability structure. Civil servants engagement is dependent on communication top-management commitment, in the absence of which they will become unmotivated.

3.10. Measurement and reporting results

As what gets measured gets managed, it is vital that both public social responsability performance and employee public social responsability engagement be actively measured and reported to executives, directors, public managers and publicly. Typically this is done in the form of an annual report.

Other human capital metrics, such as turnover, health and safety, civil servants development and diversity, for example, can be additional metrics which reveal the organsiation's social responsability commitment.

In designing public organisational social responsability reports it is important to consult public servants on what to report and what it is important that the report be received and approved by the public release departament. It is only under these conditions can the public human resources managers be assured that the organisation's public organisational social responsability performance is taken seriously.

And finally, but no less importantly, is very important to celebrate successes along the way: profile articles on the organisation's intranet, celebratory events, ensure people are congratulated on formal and also informal basis.

4. Public organizational social responsibility strategy: a new approach to innovation in public administration

At first glance, innovation and bureaucracy seem to be two opposite poles, which have no connection with each other. But viewed more closely and in depth, innovation and bureaucracy complement and influence each other without question. In circumstances where more room for changes in political and socio-economic (climate change, technological development, increasing poverty and rising unemployment), public administration can not remain the same. Ensuring citizens' interests and needs can not be achieved by ordinary activities of public organisations. It takes a series of changes that contribute to increasing professionalism in public administration, adoption of open communication and changes in management practices. We know that bureaucratic organization, from other forms of organization, has a low capacity for innovation. We could even say that the innovation process is not directly proportional to the administration. Or, rather, was not directly proportional. If until recently, bureaucratic organizational structures were characterized by rigidity in organization and management, significant technological gap and low capacity to adopt innovative ideas, a few years things have changed considerably. To date, place increasing emphasis on enhancing the innovation and modernization of public administration. Public entitites have realized that improving processes through which the public sector organization depends on innovation. Introducing innovation in public oganisations open many doors, including: maximizing use of administrative resources, improving the image and legitimacy of public institutions among citizens, development and modernization of public administration and even democratic state.

Importance of innovation for development and economic development of a state is a fact already recognized worldwide. More and more institutions and authorities to initiate and support innovative projects reliant, contributing to sustainable development of society. In this context, since 2007, Romania is carried out annually more projects in innovation and technology transfer in public administration. One of the projects is expected to have a major impact on the modernization of the Romanian public administration, in the spirit of the principles of social responsability, is initiated by the National Authority for Scientific Research: the "*public policies for sustainable innovation economy, society and environment*", comes with an innovative idea to the Romanian government, "*creating a national network of managers of innovation in central and local government*".

Innovation manager function is encountered, usually in the large corporations operating in the global market. An innovation manager is to manage the innovation process of companies and ensures that new ideas to develop into socially responsible products and services. He must plan, coordinate and evaluate the performance of specific activities of innovation. Today, more companies operating on the Romanian market have started to hire people who manages the innovation process. Human resources, financial and material, are crucial elements to ensure effective public services in the state administration. Innovation is one of the best ways to revitalize public administration. Adopting policies and socially responsible sustainable innovation in public administration and local society contribute to the development and modernization of the state. People's lives can be improved by adopting innovation in public administration.

Public organizations realize more and more the needs and benefits of incorporating sustainability in all their structures. Human resources is the key organizational partner and can take the lead or be strategic partner with other executives to work for integrate public organisational social responsability objectives into how entity gets conducted. Public human resources managers can act as translators of the organization's public social responsability commitment vertically and horizontally across departments. Everybody agree that effective human resources leadership on organisational social responsability integration requires executive and public management commitment to be successful. More and more organizational social responsability into "*all that we do*". The organisation of the future is expected to have significant transformation like public organisational social responsability is no longer becomes managed as a separate deliverable, but is part of the experience of being a civil servant in an entity that really lives its values.

Public sector organisation that do not use the expression "corporate social responsibility" are not necessarily doing any less than those that do. The challenge is for public sector bodies to identify priorities and incentives that are meaningful in the local and national context and to build on existing initiatives and capacities. There is a significant opportunity for public sector bodies in developing countries to harness current enthusiasm for "corporate social responsability" alongside key public policy goals and priorities to encourage delivery of results in both respects.

References

- Albareda, L., Lozano, J. M., Ysa T. (2007), Public Policies on Corporate Social Responsibility: The Role of Governments in Europe, Journal of Business Ethics, Volume 74, Issue 4, p. 391-40;
- Commission of the European Communities, COM 2001 366 (2001), Green Paper, *Promoting a European framework for Corporate Social Responsibility*, Bruxelles, p.3;
- Drucker, P., (1998), About manager profession, Meteor Press, București, p. 7.
- European Alliance for CSR (2008) Valuing non-financial performance: A European framework for company and investor dialogue;
- Pastor, I., Ilieş, C., (2009) The investment in the human acapital necessity of sustainable development of human resources from Romania, pp. 172-178, GEBA 2009 Globalization and Higher Education in Economics and Business Administration, Romania;
- Ilieş, C. (2009) Decision making assumptions for building corporate social responsibility, ISLS 2010 - Public and private business in Europe Global, ISLS, Third Edition, Sighisoara, Romania;
- Lockwood, N, (2004), Corporate Social Responsibility: HR's. HR Magazine;
- Lockwood, N. R. (2004), Corporate Social Responsibility: HR's Leadership Role. SHRM;
- Nelarine C., Todres, M., Shaheena Janjuha-Jivraj (2008), Woods, A. and Wallace, J., Corporate Social Responsibility and the Social Enterprise, *Journal of Business Ethics*, Volume 81, No. 2, p.5;
- Pierce, M., and Madden, K., (n.d.) *Driving Success: Human Resources and Sustainable Development.* World Business Council for Sustainable Development;
- PRIME (Principles of Responsible Management Education), The Global Impact, EABIS (European Academy of Business Academy) Ashridge (2008), *Developing* the Global Leader of Tomorrow EFMD Global Focus, Volume 03, Issue 1, 2009, accessed in June 2012;
- Report for Industry Canada, March 2009, Human Resources for Sustainability, Issue Brief and Management Checklist, , Trends and Drivers.
- Skinner, C., (2002), Links Between CSR and HR, Ethical Corporation Magazine, p.1;
- Steurer, R., (2009) The role of governments in corporate social responsibility: characterising public policies on CSR in Europe, Springer Science+Business Media, LLC;
- White, P., (2009) Building a sustainability strategy into the business, în "Corporate Governance", Vol. 9, No. 4, p. 387.

Social Enterprise and Employee Empowerment: The role of employee ownership

Sylvia Acquah¹

¹ Institute of Work Psychology, University of Sheffield, sylvia.acquah@sheffield.ac.uk

Abstract. Social enterprises are part of the so-called third sector of the economy (next to the public and private sector) and are characterised by three main components; social aim, economic goal and social ownership, with the potential to promote employee proactivity and empowerment. It could be argued that some social enterprises create an organisational setting through employee ownership (EO) where democratic decision-making is particularly prevalent with employee participation in tactical, operational and strategic decision making. Many features of social enterprises, employee ownerships in particular are consistent with the elements of empowerment from both a structural and psychological perspectives, yet, empirical evidence of such linkages are either limited or anecdotal at best. Following a review of the social enterprise and empowerment literature, two key questions emerge for better understanding the social enterprise-empowerment relation in organisations: (a) are employee-owned social enterprises more empowering?, and (b) what is the impact of EO on employee empowerment and proactivity in social enterprises? With these questions in mind, the aim of this research then is to determine the role played by the characteristics of Social Enterprise in Employee Empowerment. Exploratory qualitative study employing the Critical Incident Technique (CIT) can be used to explore the manifestation of empowerment in social enterprises, whilst questionnaire survey can be used to determine if EO contributes to empowerment in social enterprise and to compare the levels of empowerment in employee owned and non-employee owned social enterprises. A clear path for subsequent empirical studies to examine the social enterprise-empowerment relationship is therefore provided.

Keywords: Social Enterprise, Employee Empowerment, Employee Ownership, Proactivity

1 Introduction and Literature Review

Over the past 25 years, the emergence of social enterprises have had a significant impact on local, national and international initiatives aim at addressing societal issues. Social enterprises are described as a third sector organisation in the UK economy "with primarily social objectives whose surpluses are primarily reinvested for that purpose in the business or in the community, rather than being driven by the need to maximize profits for shareholders and owners" (Office of Third Sector, 2006, p. 7). A social enterprise is characterized by three main features namely (a) social aims (the purpose of the social enterprise, how it aims to benefit the society), (b) economic goal (a trading activity) and, (c) social ownership (owned and managed by the society it serves, or by the trustees or directors on behalf of the beneficiaries) (London Social Economy Taskforce, 2002). The characteristics of a social enterprise makes it an ideal public sector type organization (for addressing social needs) and a private sector type organization (as a business that generate revenue), except that most European social enterprises are distinguished by the social ownership aspect (Spear and Bidet, 2005). The employee ownership, an example of social ownership differs significantly from the main stream organisational ownership structures. Whilst most organisations are owned by investors and employees are simply labourers, employee ownership (worker cooperatives) can incorporates co-ownership by giving the employees a stake in the organisation without investing financially into the organisation.

Social ownership, particularly worker cooperatives gives employees the opportunity to collectively own the organisation. The ownership stake held by each member may differ and external parties normally do not have a stake in the organisation because the individuals exchange skills and labour for the stake; and employees participate in democratic decision-making with a one person-one vote system (Dow, 2003; Hansmann, 1996). This presents an opportunity for the employees to contribute in creating a suitable organisational climate for all which can be structurally empowering.

Empowerment has been examined from a structural and psychological perspective (Tuuli and Rowlinson, 2009). Structural empowerment is about sharing information, providing structure, providing necessary training, rewarding employees for their labour and also delegating responsibility (Kanter, 1993). These create the atmosphere or a breeding ground for psychological empowerment to take place. Sprietzer (1995) identifies four main characteristics which most psychologically empowered individual exhibit, namely: meaning (i.e. significance of employee's role to them), competence (i.e. employee's confidence in his/herself to perform), self-determination (i.e. freedom of choice, in term of how they perform their duties) and impact (i.e. the influence employee has on his/her work unit).

As discussed above, many features of social enterprises and worker/consumer cooperatives in particular are consistent with the elements of empowerment from both structural and psychological perspectives, yet, empirical evidence of such linkages are either limited or anecdotal at best. The section that follows further highlights the links between social enterprises and empowerment, whilst addressing some of the advantages of employee ownership, and outlines the key research questions that emerge from the review. A methodology for exploring the proposed research questions is also outlined.

Previous studies have linked employee ownership to motivation, proactivity, innovation, performance and job satisfaction (Michie et al. 2002; Conte and Svejnar, 1990). Social enterprises especially employee owned enterprises may therefore promote proactivity (as a result of motivation) and performance through employee empowerment and engagement. In 2010, HR Magazine for example identified employee engagement has a major role to play in driving organisations out of the current recession (Woods, 2010). Employee ownership is however a moral hazard dilemma for some organizations as it can entail the loss of control (decision making control), financial investor are normally not the sole decision makers but employees plays active role in decision making.

With specific reference to the three elements of social enterprises, the social ownership element enables ownership policies to be tailored to create an environment for equality, information sharing, joint decision making and rewarding employees, which contributes to structural empowerment. In this case, employees are more likely to assume higher levels of personal responsibility because the success of the business is their own success. Even though the majority of people associate social enterprises with charity (non-profit organization) social enterprises can be hybrid organisations operating both for-profit and as a non-profit organisation within the business model. Social enterprises like any other businesses have to generate revenue to sustain the organisation. Hence the entrepreneurs perceive a social need and adopt an innovative approach to generate revenue.

Entrepreneurs would normally possess qualities like; being able to take charge, taking risks and responsibilities and general self-efficacy (Stephan, 2010). Social entrepreneurs are also known for their pro-social motivations and behaviour which contribute to their ability to perceive social needs and incorporate social aims in the organization set-up (Stephan, 2010). Based on the course oriented structure of social enterprises, they provide employees with meaningful job which mirrors one of Sprietzer's (1995) dimensions of psychological empowerment.

From the above discussion, two key questions emerge for better understanding the social enterprise-empowerment relation in organisations; (a) Are employee-owned social enterprises more empowering? (b) What is the impact of EO on employee empowerment and proactivity in social enterprise? These questions call for a comparative research approach in order to determine whether employee owned social enterprises are more empowering in comparison to non-employee owned social enterprises and whether employee ownership within social enterprises in particular promotes employee empowerment and proactivity.

2 Proposed Methodology

These issues could better be explored through case studies (Yin, 2009) and employing systemic techniques like the Critical Incidence Technique (CIT, Flanagan, 1954) in

particular where specific episodes of empowerment and disempowerment in social enterprise context can be captured and analysed. The CIT can be used to establish which factor(s) play a significant role in employee empowerment and proactivity in employee owned and non- employee owned social enterprises. Also a questionnaire survey can subsequently be used to determine if EO contributes to empowerment, in social enterprise and compare the levels of empowerment in employee owned and non-employee owned social enterprise.

3 Conclusions

Social enterprises exhibit features that are theoretically consistent with empowerment from a structural and psychological perspective. Yet this link had hardly been tested empirically. This paper makes explicit the theoretical basis of the expectations that social enterprises, especially employee owned social enterprises, may be more empowering that their non-employee owned counterparts. A clear methodological path for subsequent empirical studies to examine the social enterprise-empowerment relationship is provided.

References

Conte, M. A., & Svejnar, J. (1990) The Performance Effects of Employee Ownership Plans. In Blinder AS (ed.) *Paying for Productivity: A Look at the Evidence*. Washington, DC: The Brookings Institution, pp. 143-172.

Dow, G. K. (2003) *Governing the Firm: Workers' Control in Theory and Practice*. Cambridge: Cambridge University Press.

Flanagan, J.C. (1954) The critical incident technique. *Psychological Bulletin*. Vol. 5, No. 4, p. 327-58.

Hansmann, H. (1996) The Ownership of Enterprise, Harvard University Press, Cambridge.

Kanter, R. M. (1993) Men and Women of the Corporation, Basic Books, New York.

London Social Economy Taskforce (2002) A Social Enterprise Business Support Strategy for London - Time to Deliver. London, Social Enterprise.

Michael, J., Oughton, C., & Bennion Y. (2002) *Employee ownership, motivation and productivity*. Birkdeck, University of London, pp. 1-33.

Office of Third Sector, (2006), *Social Enterprise Action Plan: Scaling new heights.* London: Cabinet Office of the Third Sector, pp. 1-19.

Rosen, S. (1993) Transaction Costs and Internal Labor Markets, In Williamson, O. and Winter, S. (eds), *The nature of the firms: Origins, Evolution and Development*, Oxford University press, New York, pp. 75-89.

Spear, R. & Bidet E., (2005) Social Enterprise for Work Integration in 12 European Countries: A descriptive analysis, *Annals of Public and Cooperative Economics*, Vol 76, No. 2, pp. 195-231.

Spreitzer, G. M. (1995) Psychological empowerment in the workplace: Dimensions, measurement, and validation. *Academy of Management Journal*, Vol. 38, No. 5, pp. 1442-1465.

Stephan, U. (2010) Social Entrepreneurship. In: M. Lukes and M. Laguna. Entrepreneurship: *A Psychological Approach*, Oeconomica, Prague, pp. 17-35.

Thomas, K. W. & Velthouse, B.A. (1990) Cognitive elements of empowerment. *Academy of Management Review*, Vol 15, No. 4, pp. 666-681.

Tuuli, M. M. & Rowlinson, S. (2009) Performance Consequences of Psychological Empowerment, *Journal of Construction Engineering and Management*, Vol. 135, No. 12, pp. 1334-1338.

Woods, S. (2010) Employee engagement is becoming HR's top priority, HR Magazine, London.

Yin, R. K. (2009) Case study research: Design and methods, Sage, Newbury Park.

"Determinants of the HRM policies and practices in hotel MNEs: the view from the HR department"

Giovanni Oscar Serafini¹

¹University of Sheffield Management School City College Business Administration & Economics Dept. South East European Research Centre, giserafini@seerc.org

Abstract. This study focuses on human resources management practices and policies applied in hotel multi-national enterprises (MNEs), which belong to a thriving and labor-intensive business area within the wider consumer services industry. Hence, the aim is to assess the determinants of HRM practices in different national contexts within a global luxury hotel chain that manages, franchises, owns and develops branded hotels, resorts and residential and vacation ownership properties around the world. By surveying a case study organization that operates globally within a fast growing service sector it is believed this will enable the exploration of the relationship between corporate guidelines and application of HRM policies and practices to different national contexts, thus contributing to the debate on the nature and application of HRM in the hotel industry. Important findings could occur in relation with the degree of employeremployee interdependence as well as the amount of delegation to employees. Also, as identified in the literature, useful insight will be obtained with regards to the all-important role of HR managers in the sustained implementation of HRM policies and procedures as well as the identification of the barriers that prevent the implementation of HRM in general.

Keywords: home/host country effects; role of culture; role of institutions; luxury hotel industry.

1 Introduction

This is a study of the determinants of Human Resource Management (HRM) policies in multi-national enterprises in the hospitality industry, analysing the case of a US hotel chain.

Whilst there is an already extensive literature on the determinants of HRM policy in multinationals, this study focuses specifically on hospitality where there is a limited extant literature, yet it is an industry truly global in scope. With regard to the latter, much of the literature on international HRM focuses on comparisons of a limited cross section of countries. In contrast, this case study concerns a single organization spanning over 45 countries from both developing and mature markets. Since it is recognized that there is much heterogeneity in the hospitality industry, and its focus on a premium chain entails limitations, this research allows a very much closer and direct comparison of variations in HRM policy according to locale than would otherwise be the case, within the same organization structure. In other words, it is possible to examine differences between settings within a single and mature organizational context.

More broadly, hospitality services are an essential element of a tourism product related to a destination together with transportation, events, attractions, food and drink. These, in turn, belong to a wider tourism system referred to as tourism market that interacts with the customers through transactions and impacts (Cooper et al., 2006).

Consequently, as part of tourism markets, the hospitality industry is directly affected by the sheer seasonality of demand of the tourism product: this demand fluctuation often triggered by factors outside the suppliers' control affects the management of human resources within an industry which is particularly labour intensive owing to the nature of the service processes (Reisinger, 2001). The resulting hospitalityrelated employment opportunities create jobs and the income generated has a substantial economic consequence through a "multiplier" effect, especially if a destination relies only on tourism as a main economic activity (Pender, 2005). Thus, by virtue of its link with travel and tourism, the hospitality industry has attained a global profile with an increasing importance in many national economies, thus placing local businesses in direct competition with large multinational hotel chains (Endo, 2006; Mathews, 2000).

Indeed, the hospitality industry represents a main indicator of future trends in terms of business cycle as it follows the development of markets worldwide: as an example, in coincidence with the fact that Asia, with special reference to China and India, the Pacific and Eastern Europe are forecast to become top tourism destinations by 2020 (WTO, 2001; Pine et al., 2000), the hospitality industry is growing in those regions at a noteworthy rapid pace.

According to the World Travel & Tourism Council, the travel and tourism industry is one of the most important economic sectors of the global economy, as in 2008 it employed over 225 million people around the world and generated 9.6% of global

GDP (WTTC, 2009). The World Tourism Organization (WTO, 2001) suggests that the tourism sector has grown from 25 million international tourists in 1950 to 166 million in 1970, 446 million in 1990 to reach a high of 703 million in 2002. After a decline in 2008-2009, 2010 featured a recovery in the international tourism, most notably in developing countries (WTTC, 2011).

It is estimated that over the next ten years (WTTC, 2011), the global travel and tourism economy will increase by 4.3% per year, leading its share of the global economy just above 10%. In terms of jobs creation, the WTTC estimates that 66 million jobs will be added by 2020. While in 2010 the share of total workforce employed by the travel and tourism economy was 8.1%, this figure is expected to rise by 2020 to 9.2% of total employment. Even by considering the negative effects upon tourism development caused by the 9/11 terrorist attacks, flu disease breakouts, natural calamities, at the beginning of the new millennium a WTO report had estimated that there would be 1 billion tourists in 2010, a figure that would increase by 50% within just the next 10 years. This forecast proved right, as 2010 actually demonstrated a strong recovery after the drop of 4% recorded in 2009 caused by the global economic crisis. In real terms, international tourism arrivals in 2010 reached 935 million which represented an increase by almost 7% on the previous year (WTO, 2011).

While the above themes are reviewed in subsequent sections, the potential contribution of this PhD lies in its effort to fill in a research void at two distinct levels. First, it is explicitly dedicated to the five-star global hotel industry, thus answers to the two-pronged need of more research in hotel organizations while ensuring clarity of scope within neat research boundaries. Second, following the observation by Drumm (1994) that HRM can be described as "some islands of theory in a large sea of untested hypotheses" (p. 46), this PhD case study approach is expected to push further the frontiers of knowledge. In particular, the intense case study research feature assists in answering specific research questions (Redman and Mathews, 1998) thus offering a much-needed finer understanding of developments in international HRM (Geringer et al., 2002).

Although the hospitality industry is mainly constituted of small and micro businesses (Davidson et al., 2011) analysing a global luxury hotel chain is of particular interest because of three main reasons. First of all, a hotel business of this caliber is highly capital intensive, thus plays a considerable role within foreign direct investment (FDI) in tourism of a country, which affects its economy substantially (Endo, 2006; Kundu and Contractor, 1999). Then, there are the cultural challenges linked to the management of an international company spanning over many countries that demand accommodation to peculiarities of host communities together with the imperative to maintain consistency and alignment to corporate identity and culture (Lucas and Deery, 2004). Last, owing to its labour intensive profile particularly at unskilled and semi-skilled job levels, the hospitality industry has been attempting, among others, to contain labour costs through de-skilling following the example of the fast-food industry, sometimes referred to as McDonaldization (Ritzer, 1998). Nonetheless, since the luxury end of the market addresses itself to a customer niche demanding

personalized service, it is of great interest to explore how a large and globally branded hotel chain deals with the issue of providing upscale service while attempting to contain labour costs (Davidson et al., 2011; Nankervis, 2000)

2 Literature Review

To date, literature has tended to illustrate the determinants of HR policies on MNEs through mainly cultural, institutional or globalization viewpoints and hence the ensuing sections are informed accordingly (Collings et al., 2009; Edwards and Ferner, 2002).

2.1 Home and host country effects

As any organizational expansion even at a purely national level is a challenging effort, by the same token managing organizational development of a MNE on a global scale is an extremely complex process. In fact, a MNE is asked to manage a key duality involving opposite forces affecting IHRM policies and practices implementation: on the one hand, centralization or global integration pushing for standardization and, on the other, decentralization or local responsiveness pulling towards localization (Rosenzweig, 2006). From the previous sections, it is agreed that adaptation to local context is an imperative if a transnational firm is to implement a coherent strategy throughout its subsidiaries.

The critical issue is, however, to what extent must a MNE adapt its country-of-origin practices to meet its host-country needs while it is demonstrated that home-country influences exert strong pressures on management systems, the culture of the organization as well as the application of human resource management processes (Lau and Ngo, 1996; Rosenzweig and Nohria, 1994). In particular, Taylor et al. (1996) indicate that MNEs, by drawing on their home-country competitive advantage, tend to implement homogeneous HRM systems in their subsidiaries overseas. The researchers show that this attempt is especially successful when the gap between the parent and host countries in terms of cultural and legal distance is limited.

According to Cyr and Schneider (1996), international joint ventures (IJVs) in transition economies of the former Eastern Bloc represent an interesting case in point whereby successful performance is the result of effective co-operation between locals and foreign nationals. As key prerequisites, the researchers indicate that "trust and mutual respect" (p. 219) are instrumental for the attainment of an enduring employee commitment, which, in turn enables the organization to learn and grow while relying on locally sound communication, training, staffing, appraisal and reward systems. Furthermore, in their recent research of Eastern European transition economies, Jindra et al. (2009) highlight that beneficial technological diffusion and spillovers towards the development of host-country economies are a direct consequence of a MNE subsidiary role and its technological competence. Again in this specific geographic area, which represents an ideal laboratory to test HRM implementation due to its Communist past, Lewis (2008) indicates that one of the most precious contribution of MNEs to their subsidiaries and society at large are the training and developmental opportunities offered to the local workforce while instilling the meanings of productivity and quality. As Kostova (1999) argues, the transfer of organizational practices within a MNE is effective provided that the social, organizational, and relational dimension are fulfilled.

Equally important, Ferner and Edwards (1995) indicate that MNEs should also be considered as structures of power whereby operational outcomes are the effect of a dynamic interplay among resource possession and exchange, formal authority and vested interests. Within this viewpoint, the transfer of best practices is greatly facilitated by coercive comparisons which allow for the exertion of pressures on local management and workforce to align accordingly (Sisson et al., 2003). Nonetheless, Martin and Beaumont (1998) contend that MNEs should exert a viable pressure on their subsidiaries following the example of a model company such as ABB. More in detail, they remark that internal benchmarking should, on the one hand, consider the local context in terms of culture and institutions and, on the other hand, the capability and motives of local management to apply best practices.

From a different angle and in contrast with Minbaeva et al (2003), Szulanski (1996) indicates that the diffusion of best practice within an organization has to be fostered by eliminating "internal stickiness" (p. 29) caused by elements beyond purely motivational aspects. Indeed, his research leads to the conclusion that barriers to knowledge transfer are caused, firstly, by the inability of the workforce to identify, appreciate and implement new knowledge; then, lack of depth of knowledge; and, last, by ineffective communication and poor interpersonal relationships. As Edwards (1998) contends, if there is prevalence of environmental elements favouring synergies and appropriate organizational "maturity" conditions in a global MNE, there may be cases of "reverse diffusion" (p. 696) according to which practices are even transferred from the international subsidiaries to other subsidiaries and domestic plants of the organization.

2.2 Institutions and home country effects

Ferner and Varul (2000) argue that IHRM is critically affected by the country of origin model of personnel management, which is conditioned by factors within the national business system. Their research outcome is the result of an analysis of the highly structured German institutional framework determining a considerably reactive and administration-centred personnel management approach pressing against the strategic adaptation of a wider IHRM implementation along the lines of the 'Anglo-Saxon' model.

Parent country (ie, country-of-origin) effects represent a category of elements exerting pressures on the way multinationals operate, their organizational culture and, ultimately, their approach to HRM. Thus, MNEs from different home countries apply IHRM in different ways for the very reason that, according to Ferner (1997):

"even the most global of companies remain deeply rooted in the national business systems of their country of origin.(...) Even where the home base does not account for the bulk of sales, operations and employment, the home nation is almost always the primary locus of ownership and control" (p. 19).

As illustrated by Chang et al. (2007) the pressures exerted onto a MNE in order to be aligned with home country institutional environments represent "push forces" (p. 405) and are the result of the MNE degree of embeddedness in its business system (Ferner, 1997). The home country effect can take many forms as demonstrated in numerous studies comparing practices in MNEs from different countries. In particular, nationality of ownership is a critical determinant of MNE business code of conduct and management practices following the comparison of US MNEs as opposed to Japanese and EU MNEs, for instance. While the former tend to be centralized and standardized in the management of HR, Japanese MNEs are more keen to adapt to the local settings (Ferner, 1997). Also, the fact that MNEs mostly gravitate around their home country (Hejazi, 2007) is demonstrated in their concentration of assets, sales generated as well as the marked tendency to fill key positions at the executive level with parent country nationals Edwards et al. (2007). The latter, in fact, plays a key role in the transfer process of knowledge and management practices as well as the determination of industrial relations policies and practice from headquarters to subsidiaries Edwards et al. (2007).

2.3 Organizational isomorphism

The implementation of IHRM by a MNE is not a straightforward process, but rather reflects the modus vivendi of subsidiaries and their dynamic interplay vis-à-vis the organization headquarters. In fact, Rosenzweig and Nohria (1994) in studying human resource management practices in 249 U.S. affiliates of foreign-based MNEs from Canada, Japan and Europe demonstrate that, broadly, affiliate HRM practices are strongly aligned with local practices while, however, differing in particular practices. The extent of "local isomorphism" (p. 241) as the authors put it, is determined by the founding method, the reliance on local inputs, the ratio of expatriates within the workforce of the subsidiary, the closeness of communication with headquarters and the need for internal organizational consistency. According to DiMaggio and Powell (1983, p. 150) there are three main forms of isomorphic outcomes, each determined by particular processes: firstly, "coercive isomorphism" relates to the pressures onto an organization as occurs through the imposition of law, for instance; secondly, "mimetic isomorphism" indicates efforts of an organization to copy successful practices adopted by other organizations; and, last, "normative isomorphism" describes the adjustment of an organization to what is regarded as appropriate in a particular environment (Kostova and Roth, 2002; Haveman, 1993)

Notwithstanding, Jackson and Schuler (1995) highlight that local environment and culture have a key role in the application of HRM systems, policies and practices. On this same line, a study by Beechler and Yang (1994) on Japanese subsidiaries operating in the USA shows that local circumstances decisively influence a MNE's

ability to transfer parent-country practices abroad. As a result of particular local constraints, businesses are mostly unable to accurately replicate home countries' practices in their international subsidiaries and thus opt for a locally isomorphic approach Rosenzweig and Singh (1991). Indeed, even between similar national business systems such as the American and the British, the transfer of HR practices from the US parent company to the UK subsidiary is impeded by local stakeholders who prevent a MNE to fully implement a corporate diversity policy, as shown in a case analysis by Ferner et al. (2005).

2.4 HRM in multinational enterprises

To understand the nature of HRM in a hotel MNE requires initially a review of the general literature on the determinants of HRM in MNEs and then move on to the existing literature on hotels. Thus, for the purposes of the research, the landmark Dunning and Lundan (2008) definition of a MNE is adopted, which states that:

"A multinational or transnational enterprise is an enterprise that engages in foreign direct investment (FDI) and owns or, in some way, controls valueadded activities in more than one country." (p.3)

Perlmutter (1969) identifies four major approaches to the management of overseas subsidiaries. First, the ethnocentric approach maintains that values, culture and strategic decisions are defined by the expectations of the mother company, thus offering very limited power or autonomy to the subsidiaries overseas. Since these are predominantly managed and controlled by expatriates or former staff at the headquarters, locals have minimal input on policies and procedures as communication occurs through directives elaborated and issued by HQ.

Second, the polycentric approach is characterized by a softer pattern according to which each international subsidiary is considered as an autonomous business unit, controlled and managed by local managers. Indeed, while key decision making related to financial and strategic issues is a prerogative of headquarters whereby the key positions at that level are exclusively covered by people from the parent company, local managers are given higher profile tasks which reflect their deeper understanding of local requirements in the areas of marketing, production and human resource management (HRM). Consequently, subsidiaries are allowed a greater degree of autonomy enabling the maintenance of policies and practices pertaining to employee relations which are aligned with the culture and regulatory framework of the host country.

Then, the regiocentric approach is characterized by further devolvement by HQ according to which decision making and workforce management is implemented on a regional or geographical basis, but top positions are still held by nationals of the parent company's country. Nevertheless, regional managers maintain enhanced power and autonomy even if these are limited within the confines of the region where they operate.

Last, the geocentric model is proper of increasingly globalized companies whereby the management of overseas subsidiaries occurs through a combination of both home country and parent company managers. Within this perspective, it is vital for organizations to nurture a corporate environment characterized by a seamless exchange of ideas, values, information and best practices. Nonetheless, as Edwards et al. contend (2007), in practice there are often insurmountable challenges to transferring in a standardized fashion policies, practices and procedures of a MNE. Thus, the implementation of such an homogenizing approach risks to become an impossible task due to the fact that operating procedures may be transferred in full, partially, or not at all depending on the host country business system.

As per Perlmutter's model, for both last approaches there is no decisive influence of any particular culture, either of the home or of the host country. While it is granted that local legal requirements are fulfilled in the different locations where the MNE operates, such an organization tends to develop overarching regional or global approaches, policies and practices shared internally. Still, following the categorization by Chang et al. (2007), pull forces are all those tensions onto a MNE in order to conform to the institutional environment of the host country (ie, country of domicile). In particular, local business systems may hamper the MNEs in their efforts to replicate home country practices throughout their overseas subsidiaries. Both host country institutions and culture represent the main obstacles against the standardization of practices across MNEs (Brewster et al., 2008) and hence pulling towards localization (Myloni et al., 2007; Beechler and Yang, 1994).

Beyond the above model illustrating the four different approaches to managing subsidiaries of multinationals, there are two additional approaches, namely diffuse diversity and mix and match patterns. In the incoherent process of diffuse diversity, Brewster et al. (2008) identify "a huge range of different practices within and across national economies" (p. 333) whereby no clear approach is visible. Instead, in the latter mix and match pattern there is a combined effect whereby some features of practice are in line with what is commonly encountered in the parent country and some in the subsidiary country/host country (Buckley and Ghauri, 2004). In particular, Buckley and Ghauri (2004) contend that even if the global economy is converging through the efforts of MNEs, still there are "conflicts between markets and national policies" (p. 82) that hamper global integration. Beyond such conflicts, location and ownership strategies of MNEs to co-ordinate global activities are confronted with the issue of distance at cultural, institutional and geographical level (Beugelsdijk et al., 2010). As illustrated by Johnson and Vanetti (2005) operational inconsistencies and differentiations from home country standards can be minimized across MNE subsidiaries through tight control by means of extensive codification as occurs in the case of hotel franchising agreements (Contractor and Kundu, 1998), which still offer some flexibility to franchisees and owners to the extent that minimum requirements are respected.

However, while the local environment may affect the consistent implementation of HRM policies and procedures emanating from the MNE headquarters (Muller, 1999),

Lane (2006) argues that institutions, while posing boundaries, must not be seen as "totally constraining actors" (p. 84). In fact, Ferner et al. (2001) observe that, as in the case of Spain which features a highly regulated labour market, MNEs operating there have managed to attain a degree of flexibility allowing for the preservation of elements proper of their home style.

Last, given the considerable importance of a MNE subsidiary to the economy of the host country since it represents Foreign Direct Investment, Schnitzer (1999) indicates that "an implicit or self-enforcing contract is sustained by a MNE's threat to withdraw" (p. 1131). Within this framework, the risk for a host country to see a MNE withdraw its operations to a more convenient overseas destination allows the subsidiary to gain the necessary freedom to replicate HRM policies and procedures emanating from headquarters. By the same token, MNEs may capitalize on such threat to ensure their policies are diffused throughout the international subsidiaries.

Relating the above theoretical account to hospitality organizations requires considering the wider tourism industry and its stage of globalization because the former is closely intertwined with the latter (Hjalager, 2007). Even if the hotel industry figures among the most global in the service sector (Littlejohn, 1997) and despite the tendency of markets towards globalization thus pushing hospitality organizations towards the standardization of both products and services, nonetheless there is the imperative to adjust to local demands peculiar to national markets and environments.

The different cultural orientation by organizations illustrated by Permutter (1969) reflects on the international hospitality industry at two main levels. As previously mentioned, an ethnocentric orientation assumes that practices developed in the parent country of a MNE are superior to others or not requiring modifications owing to similarities with other countries Geppert and Clark (2003). Thus, firstly, recruitment strategies aimed particularly at manning worldwide operations would reveal the kind of international management development programs implemented as well as the choice of international managers to be developed. If ethnocentrism and polycentrism represent the extremes of a continuum whereby the former supports the use of home country expatriates in subsidiaries overseas and the latter advocates the employment of local nationals, a geocentric approach stands in the middle as it maintains the use of a mixture of home, host and third country nationals (D'Annunzio-Green, 1997). Consequently, hospitality organizations implementing an ethnocentric approach develop strategies in the parent country and rely on expatriated nationals to manage local staff in host countries in order to determine the right implementation of organization structure and processes (Edstroem and Galbraith, 1977).

From an opposite standpoint, polycentric organizations maintain a market-driven strategy based on divergence since they assume that products and services need to be customized according to customer segments: thus, emphasis is shifted towards divisional or, even, unit-level managers and the practices they establish (Roper et al., 1997). On the other hand, a geocentric business approach tries to strike a balance between ethnocentrism and polycentrism because it is world oriented and

characterized by a supra national framework (Perlmutter, 1969). Although this is the case of leading world hotel chains such as Holiday Inn and Hilton according to Nickson (1997), still their worldwide business expansion in their early days by their founders was characterized by the "dissemination of American management and operating techniques" together with the attempt "to diffuse "best practice" techniques of modern hotel management" (p. 187). Last, hospitality organizations applying a regiocentric approach tend to mix geocentrism with polycentrism in an effort to satisfy the opposed needs of standardization and customization.

A shining example of this formulation is the implementation by Hilton of the 'Wa No Kutsurogi' program (Teare, 1993) to cater to the specific needs of the Japanese traveller thus addressing the demand for customization within a multicultural context (Pullman et al., 2001). Indeed, market drivers influence the strategic decisions of consumer-driven hospitality businesses to cater to the needs and expectations of both local and international customers. Efforts to provide unique hospitality services to capitalize on market potential thus push hospitality organizations towards local responsiveness thus fragmenting the competitive context (Bender et al., 2008; Whitla et al., 2007). Nonetheless, hospitality MNEs could be forced to adjust to local realities owing to the critical role of other factors such as investment environment, political stability, nature of institutions as well as employment law of the host countries where MNE subsidiaries are set to operate (Guillet et al., 2011; Thite et al., 2011; Tuunanen et al., 2011; Royle and Ortiz, 2009).

On the other hand, hospitality MNEs need to focus on decision making processes which are globally oriented in order to maintain their brand power and penetration across geographic locations, an effort which is also challenged by the very nature of hospitality business model expansion whenever it separates hotel ownership from management. In fact, where hotels are not owned and operate either through franchise or management contract agreements, the local owning company tends to put their interests above those of the hotel chain thus potentially hampering any attempts to effectively consolidate an integrated as well as centralized global system (Whitla et al., 2007; Brotherton and Adler, 1999).

2.5 Etics and emics: recognizing culture in varying contexts

From a wider perspective, researchers have identified two broad conceptualization categories according to which host cultures can be explained: the etic and the emic approaches. The etic approach aims, as Morris et al. (1999) argue, to "describe phenomena in constructs that apply across cultures" (p.782) whereby it attempts to illustrate cultural differences in terms of some universal standards applied to human behavior. Consequently, the work of Hofstede (1980) evidently belongs to the etic proposition (Brewster et al., 2008; Morris et al., 1999).

Alternatively, the emic framework maintains that culture is inextricably linked with the broader social system thus tends to examine constructs from a viewpoint within specific cultures (Morris et al., 1999). To better illustrate these differing viewpoints,

Gudykunst (1997) makes reference to an earlier exemplification by Pike, the linguist who first introduced the concepts of emics and etics (Peterson and Pike, 2002; Berry, 1999; Feleppa, 1986) whereby the differentiation between etics and emics approaches corresponds to the "discussion of phonetics (vocal utterances that are universal) and phonemics (culturally specific vocal utterances)" (p.329). As a result, emics and etics are not dichotomous, but rather need to be seen as complementing each other in a symbiotic fashion whereby, as Berry puts it (1999) their interaction offers the opportunity to learn about "human development and behavior within and across cultures" (p. 170).

By drawing on the different perspectives emics and etics approaches take, Maheswaran and Sharon (2000) indicate that research with an emic orientation "favors within-culture investigation, arguing that theorizing is culture-specific and should, therefore, be inductive" (p. 60). Emics represent a key issue in the current cultural psychology debate and, specifically, in relation with cross-cultural research. Thus, in applied terms, emics are key in the study of a wide variety of organizational study topics particularly as contrasted with etics (Peterson and Pike, 2002).

In the final analysis, emics approach is critical in contributing depth to explanations, be it a study on leadership in cross-cultural contexts (Dickson et al., 2003) or related to cultures in a globalized world (Bhawuk, 2008) or, even on management accounting (Lukka and Modell, 2010). As a further example, occupational therapy is a field that requires emics in order to explore issues from the insider's view and provide due relevance to the account (Reid et al., 2005).

Challenges in both these approaches is that while the etic approach may lead to arbitrariness in the selection of predetermined standards ending in stereotypes, the emic approach may restrain any initiative towards the management of people across nations. Since human behavior may feature both emic and etic characteristics, emics and etics are not necessarily incompatible but rather complement each other (Berry, 1999).

Nonetheless, since both emics and etics concede that culture is taken for granted as it is deeply embedded within societies (Brewster et al., 2008), it is a challenging task to transform cultures towards the enhancement of their social capital because social trust cannot be established deliberately as well as familiar norms and rules are difficult to be modified (Fukuyama, 1995). As a result, in the context of organizations, this translates into culture offering to employees the opportunity of making sense of their work role (Harris, 1994) while providing the framework towards higher performance levels and organizational development (Brown, 1992).

2.6 The role of culture

Since HRM is concerned with an organization's human asset, the role of culture is capital in order to understand the way employees behave and thus perform.

According to Hofstede (1984), a pioneering researcher in the field, culture is mainly built around values and is defined as being:

"the collective programming of the mind which distinguishes the members of one human group from another" (p.21).

In his landmark research, Hofstede (1980) defined that HRM policies and practices may not be valid across different countries because of divergence from local laws, customs and cultures. Equally, cultural differences could affect how the HR function is deployed across borders by global organizations. In 1980, he was able to study how cultures differentiate from country to country by examining an impressive amount of around 117,000 responses to employee attitude surveys held between 1967 and 1973 administered at the subsidiaries of a MNE (IBM) in 71 countries (Hofstede and McCrae, 2004). Thus, while he initially identified four cross-cultural dimensions, he later (Hofstede, 1993) included a fifth dimension that led him to forge a model aiming at representing a universally valid taxonomy of traits. In particular, the five cross-cultural dimensions identified by Hofstede are Uncertainty Avoidance, Masculinity-Femininity, Individualism-Collectivism, Power Distance, and Short-Long Term Orientation.

First, Uncertainty Avoidance relates to the extent people are at ease with ambiguity. Personalities with high levels of Uncertainty Avoidance confer great value to stability and certainty, and tend to appreciate methodologies that heighten a sense of security in unclear situations (e.g., detailed work plans). Individuals who score low in this dimension are more comfortable with risk and are more prone to move forward without having the full picture of what is lying ahead.

Then, Masculinity-Femininity illustrates the degree that personalities show masculine (e.g., independence, dominance) or feminine values and behaviours (e.g., interdependence, empathy, openness). High masculine cultures feature clearly differentiated gender roles, independent performance, achievement, and ambition, while feminine cultures are characterized by equal gender roles, quality of life, and helping others.

Thirdly, the author considers Individualism-Collectivism as being the dimension that refers to the degree according to which individuals emphasize their personal needs and wants versus the needs of the group. Thus, responsibility and achievement by individuals belong to cultures that tend towards the individualism side of this cultural continuum, while cultures that are closer to collectivism focus on group efforts, teamwork, and group membership.

Following is the fourth dimension that deals with Power Distance, which reflects the degree that people accept differences among individuals as sensible and expected. Therefore, cultures featuring high Power Distance determine workplace environments whereby subordinates and their superiors constitute two distinct groups separated by a considerable status gap. To the other extreme, people belonging to low power

distance cultures experience fewer status barriers in their reporting relationships because superiors are more open.

Last, the fifth dimension is represented by the Long-Short Term Orientation, whereby Short-Term Orientation emphasizes the present or the past, honours customs and tradition, and utilizes resources to meet current needs. On the other hand, environments featuring a Long-Term Orientation are geared towards the future by saving resources to meet future emergencies.

Based on the vast amount of data collected, Hofstede (1993; 1980) has been able to plot many countries along the five cross-cultural dimensions according to the scores recorded. This resulted in the production of extremely valuable information material featuring national profiles that are of both theoretical and practical benefit to IHRM.

Consequently, any organization planning its operational expansion on a global scale may attain valuable understanding of successful management approaches and practices in overseas locations by referring to the country profiles defined by Hofstede. For example, Power Distance scores high in Latin, Asian and African countries, while it is reduced in Germanic ones. As for Uncertainty Avoidance, Latin countries, Japan and German-speaking countries rank high as opposed to Anglo, Nordic and Chinese-culture countries. Next, Western countries feature Individualism markedly while Collectivism is widespread in less developed and Eastern countries, with Japan lying in the middle of the two extremes. Last, Masculinity is overwhelming in Japan and in some European countries such as Germany, Austria and Switzerland, and moderately high in the UK. On the opposite, Masculinity is low in the Nordic countries and in the Netherlands, while being moderately low in certain European and Asian countries like Spain, France and Thailand (Hofstede and McCrae, 2004).

These cultural differences manifest themselves at multiple organizational levels within a MNE: according to Glover and Wilkinson (2007), management need to come to terms with so-called tensions that relate to both the external and internal environments of an organization when trying to apply strategies to a subsidiary of a foreign MNE. Hence, the importance of a balanced approach to management that, by ensuring fit between centralization and decentralization (Ferner et al., 2004) as well as harmony among practices and national culture, attains higher financial performance and better strategy implementation than otherwise (Bae et al., 1998). Interestingly, however, cultural differences are apparent not only in cross-border mergers and acquisitions but also within a same country with identical economic market type thus compelling HRM to take initiatives aiming at recognizing highly localized peculiarities (Aguilera and Dencker, 2004).

Starting with the cultural orientated approach, its critical benefit is that, drawing from the construct of culture, it attempts to explain or foresee human behavior where, in practical contexts, the mere consideration of economic, political or institutional factors is inadequate (Hofstede, 2002). As Chapman maintains (1996), the virtue of the cultural approach is that it has sparked much fruitful debate on the effects of

culture on business and management. More recently, Amable (2008) recognizes the undeniably mediating effect of culture on business systems and organizational capabilities through its influence, for instance, on "behavioural norms" (p. 781) or the "perception of fairness" (p. 783).

2.7 Cultural orientated v. institutional approaches

A most recent strand of research has embraced the institutional perspective because scholars recognize its ability to integrate social complexities by taking into account all elements embedded in the context, such as economic rules, social actors, institutions, organizations as well as behaviours (Djelic, 2010; Jackson, 2010; Whitley, 2006; Boyer, 2005). In particular, Whitley contends that the institutional approach allows for a fine-grain examination linking systems with business realities (Whitley, 2006). This rational approach not only relies on objective observations and concrete data, but also, most importantly, facilitates industrial comparative analyses (Jackson, 2010, p. 81) as in identifying the role of institutions upon MNEs (Whitley, 2006). Consequently, since this PhD research focuses on a comparative analysis within a MNE, the institutional approach is adopted.

Theories of globalization emphasizing the application of worldwide practices encompass, beyond the ethnocentric approach previously illustrated, also convergence and homogeneity frameworks (Brewster et al., 2008). More in detail, the latter contend that owing to the fact that the world is getting nationless and borderless (Ohmae, 1996,1990) over time economies will become increasingly more similar thus converging towards a mainly Anglo-American free market model. As a result, Chow (2004) observes that increasing pressures originated by industrialization and globalization forces push towards the application of similar HRM practices.

2.8 Differences in HRM implementation: hospitality MNEs v. local hotels

Although rising, the number of multinational organizations yet represent the sheer minority of all businesses operating in the hospitality industry: in fact, local small and medium sized hotel enterprises (SMEs) prevail all the world over as documented by Baum (1999) according to whom 90% of hospitality businesses globally belong to the SME type. As a result of their intrinsic lack of professional and business skills, insufficiency of funds and unstable business performance, most SMEs are evidently disadvantaged when it comes to competing on human resources management grounds with larger organizations (Baum, 1999).

Thus, the hospitality industry is theatre to a most striking contradiction: despite the fact that theoretical propositions and empirical evidences clearly indicate that HRM effectiveness is key to service quality, customer satisfaction, business sustainability and profitability as well as competitive advantage, still the reality of many local hotel organizations in terms of workforce management practices and employment conditions often does not reveal the importance of sound HRM (Cho et al., 2006).

More worrying is the fact that in practice the role of HRM within such organizations is given secondary priority compared to other business functions such as finance and sales and marketing (McEvoy, 1984; Rutherford and O'Fallon, 2007; Haynes and Fryer, 2000; Kelliher and Johnson, 1997). This leads to questioning the credibility of the fundamental assumption made and expressed by hotel management according to which employees are the most important asset and key to success in relation to service quality, consumer satisfaction as well as business performance (Losekoot et al., 2001; Maher, 1993).

On the other hand, hospitality MNEs with an international or, even, global reach have made remarkable efforts to narrow the gap between words and deeds and upgrade the profile of the industry by realizing the long-term benefit of sound HRM practices implementation (Bohdanowicz and Zientara, 2009; Rowley and Purcell, 2001). As Baum (2006) maintains, it is vital for hospitality organizations to develop their workforce so as to render them "more flexible and adaptive to constant change" (p. 133) at both operational and managerial levels. Thus, Ford and Heaton (2001) contend that, due to the considerable intangible aspect of the service experience (Lashley, 1998), the hospitality industry has managed to develop unique competencies in service provision that businesses in any other industrial sectors may consider worth implementing. This because while the hotel industry is particularly capital- and labor-intensive, its logistics and supply chain processes can be as sophisticated as in manufacturing businesses (Dimou et al., 2003) contrarily to other sectors of the services industry, such as marketing and consulting.

3 Data/Methodology

The company under consideration is a global deluxe hotel chain that, for the purpose of safeguarding anonymity, will be referred to as HotelCo. This offers a unique research opportunity to understand the influence exerted by the MNE centre towards the periphery in an industry where MNEs are more spatially bound than many other employers. In fact, the peculiarity of international hotel chains is that they cannot, unlike manufacturing and some services such as call centres, move to the locations of lowest possible cost in a race-to-the-bottom strategic choice as previously explored (Contractor et al., 2010; Wright et al., 2005). Rather, they have to base their decisions primarily on consumer demand, resulting in hotel chains being present in some extremely high wage and heavily regulated contexts (Contrepois and Jefferys, 2010; Knox, 2010; McDowell et al., 2007). This makes hotels a very interesting case because they cannot simply exit to avoid regulation as a blanket policy since they need to be where customers are (albeit that they may choose to leave a specific market if the costs of labour, for example, are so great as to make for uncompetitiveness), but rather optimize their strategy according to the particular context and competition settings (Timo and Davidson, 2005).

Thus, the case study organization is a diversified global US hospitality concern with more than fifty years of history, acting as the owner-manager, franchiser and

management company of international deluxe hotels, resorts, and mixed-use real estate developments. The HotelCo global mission is to provide authentic hospitality and its goal is to be recognized as the most preferred brand in each market segment it operates while being faithful to a set of core values featured in its culture. In particular, as of March 31, 2012, HotelCo's worldwide portfolio consisted of 488 properties grouped under the EAME (Europe-Africa-Middle East), ASPAC (Asia-Pacific) and LATAM (Latin America) divisions which are coordinated by divisional offices, respectively.

Therefore, through the application of a cross-national comparative study, the aim of this PhD is to examine the determinants of HR practices within a MNE operating in the hospitality industry and attain new insights into the relative impact of host and home country effects. With the premise that the research aims at surveying a leading global hospitality company operating in 45 countries, the consideration of the national archetypal business systems as enunciated by Whitley (1999) is instrumental in order to group national contexts according to their distinctive type of coordination and control systems. In particular, Whitley's model (1999) includes: fragmented, coordinated-industrial-district, compartmentalized (i.e. 'Anglo-American model'), state-organized, collaborative ('the Rhineland model') and coordinated ('the Japanese model') business systems.

In the final analysis, using a case study organization that operates globally within the fast growing service sector it is believed it will enable the exploration of the relationship between corporate guidelines and application of HRM policies and practices to different national contexts, thus contributing to the debate on the nature and application of HRM in the hotel industry.

4 Empirical Analysis

Since the overall aim of this PhD research is to assess the nature and determinants of HR practices of a US hotel MNE both in its home and foreign subsidiaries from the viewpoint of HR executives, it is necessary to attain the following research objectives:

i) to identify, within the context of the luxury hospitality industry, the degree of autonomy subsidiary HR executives have in setting their HR policies;

ii) to explain the extent to which subsidiary level HR executives implement HR policies and procedures in line with those encountered in the country of operation;

A quantitative survey method will be used to examine the approaches and implementation of HRM policies and practices in a global luxury hotel chain. As Saunders et al.(2007) point out, the quantitative survey "is a popular and common strategy in business and management research", and surveys enable "easy comparisons and appear to be authoritative" (p. 92). Thus, the survey is a useful tool

for analyzing and assessing HR practices in a systematic way that allows the researcher to gather similar data and draw useful conclusions following comparisons.

Information is sought in order to assess the extent of formal corporate HRM strategy application by subsidiaries, drawing from the input via questionnaire administration to HR executives operating at home and overseas subsidiary level in properties located in EAME (Europe-Africa-Middle East) and the US encompassing a sample of around 50 respondents. Owing to the particular organizational structure of the MNE organization, the HR function is arranged accordingly. Thus, field HR executives operating at hotel unit (ie. subsidiary) level report to the Area HR Director, who, in turn, reports to the respective divisional office. HR Divisional Offices then report to the HR headquarters located in the US. The fact that participants to the survey are Human Resources executives should provide a unique insight into the process directly from the management representatives in charge of implementing HRM policies and practices at hotel unit level. The wealth of information obtained allows for direct comparison across subsidiaries (Pudelko and Harzing, 2007). Additionally, data gathering is then enriched through semi-structured follow-up interviews to HR executives located at the divisional and headquarter offices of the HotelCo.

5 Conclusions

This article aims at laying the theoretical grounds to an ensuing research work in the field of HRM applied to the global luxury hotel industry. The review of literature leads to the conclusion that business and context specificities including host country effects, the role of culture and institutions plus globalization pressures affect the application of HRM in practice across countries where a MNE operates.

Being the hotel group a global organization, the research will offer useful insight on how people management is occurring in different countries and contexts thus allowing the interpretation of results across locations and economies.

References

- Aguilera, R. V. and Dencker, J. C. (2004) The role of human resource management in cross-border mergers and acquisitions. *The International Journal of Human Resource Management*, 15 (8), p. 1355 - 1370.
- Amable, B. (2008) Richard Whitley business systems and organizational capabilities: the institutional structuring of competitive competences: the heterogeneity of competing capitalisms and business systems. *Socio-Economic Review*, 6 (4), p. 771-775.
- Bae, J., et al. (1998) Variations in human resource management in Asian countries: MNC home-country and host-country effects. *The International Journal of Human Resource Management*, 9 (4), p. 653 - 670.
- Baum, T. (1999) Human resource management in tourism's small business sector: policy dimensions. In: Lee-Ross, D. (ed.) HRM in tourism and hospitality: international perspectives on small to medium sized enterprises. London, Cassell, p. 3-16.
- Baum, T. (2006) The future of work and employment in tourism. In: Buhalis, D. and Costa, C. (eds.) Tourism management dynamics: trends, management and tools. Oxford, Elsevier Butterworth-Heinemann, p. 130-136.
- Beechler, S. and Yang, J. Z. (1994) The transfer of Japanese-style management to American subsidiaries: contingencies, constraints, and competencies. *Journal of International Business Studies*, 25 (3), p. 467-491.
- Bender, B., et al. (2008) An examination of strategic drivers impacting U.S. multinational lodging corporations. *International Journal of Hospitality & Tourism Administration*, 9 (3), p. 219-243.
- Berry, J. W. (1999) Emics and etics: a symbiotic conception. *Culture & Psychology*, 5 (2), p. 165-171.
- Beugelsdijk, S., et al. (2010) Introduction: Place, space and organization economic geography and the multinational enterprise. *Journal of Economic Geography*, 10 (4), p. 485-493.
- Bhawuk, D. P. S. (2008) Globalization and indigenous cultures: homogenization or differentiation? *International Journal of Intercultural Relations*, 32 (4), p. 305-317.
- Bohdanowicz, P. and Zientara, P. (2009) Hotel companies' contribution to improving the quality of life of local communities and the well-being of their employees. *Tourism and Hospitality Research* 9(2), p. 147-158.
- Boyer, R. (2005) How and why capitalisms differ. *Economy & Society*, 34 (4), p. 509-557.
- Brewster, C., et al. (2008) Similarity, isomorphism or duality? Recent survey evidence on the Human Resource Management policies of multinational corporations. *British Journal of Management*, 19 (4), p. 320-342.

- Brotherton, B. and Adler, G. (1999) An integrative approach to enhancing customer value and corporate performance in the international hotel industry. *Hospitality Management*, 18 (261-272).
- Brown, A. (1992) Organizational culture: the key to effective leadership and organizational development. *Leadership & Organization Development Journal*, 13 (2), p. 3-6.
- Buckley, P. J. and Ghauri, P. N. (2004) Globalisation, economic geography and the strategy of multinational enterprises. *Journal of International Business Studies*, 35 (2), p. 81-98.
- Chang, Y. Y., et al. (2007) HRM strategies and MNCs from emerging economies in the UK. *European Business Review*, 19 (5), p. 404-419.
- Chapman, M. (1996) Social anthropology, business studies, and cultural issues. (cover story). *International Studies of Management & Organization*, 26 (4), p. 3-29.
- Cho, S., et al. (2006) Measuring the impact of human resource management practices on hospitality firms' performances. *International Journal of Hospitality Management*, 25 (2), p. 262-277.
- Chow, I. H.-S. (2004) The impact of institutional context on human resource management in three Chinese societies. *Employee Relations*, 26 (6), p. 626 642.
- Collings, D. G., et al. (2009) International human resource management. *In*: Collings, D. G. and Wood, G. (eds.) *Human Resource Management: a critical approach*. New York, Routledge, p. 296-312.
- Contractor, F. J., et al. (2010) Reconceptualizing the firm in a world of outsourcing and offshoring: the organizational and geographical relocation of high-value company functions. *Journal of Management Studies*, 47 (8), p. 1417-1433.
- Contractor, F. J. and Kundu, S. K. (1998) Franchising versus company-run operations: modal choice in the global hotel sector. *Journal of International Marketing*, 6 (2), p. 28-53.
- Contrepois, S. and Jefferys, S. (2010) European Works Councils in Central and Eastern Europe: varieties of institution building among French service sector multinationals. *Industrial Relations Journal*, 41 (6), p. 584-602.
- Cooper, C., et al. (2006) New and emerging markets. *In*: Buhalis, D. and Costa, C. (eds.) *Tourism business frontiers: consumers, products and industry*. Oxford, Elsevier Butterworth-Heinemann, p. 19-29.
- Cyr, D. J. and Schneider, S. C. (1996) Implications for learning: Human Resource Management in East-West joint ventures. *Organization Studies*, 17 (2), p. 207-226.
- D'Annunzio-Green, N. (1997) Developing international managers in the hospitality industry. *International Journal of Contemporary Hospitality Management*, 9 (5/6), p. 199-208.
- Davidson, M. C. G., et al. (2011) Hospitality HRM: past, present and the future. International Journal of Contemporary Hospitality Management Volume, 23 (4), p. 498-516.
- Dickson, M. W., et al. (2003) Research on leadership in a cross-cultural context: Making progress, and raising new questions. *The Leadership Quarterly*, 14 (6), p. 729-768.

- DiMaggio, P. J. and Powell, W. W. (1983) The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48 (2), p. 147-160
- Dimou, I., et al. (2003) Corporate development of international hotel firms: a transaction cost economics and agency theory approach. *International Journal of Tourism Research*, 5 (2), p. 147-149.
- Djelic, M.-L. (2010) Institutional perspectives working towards coherence or irreconcilable diversity? *In*: Morgan, G., et al. (eds.) *The Oxford handbook of comparative institutional analysis*. Oxford, Oxford University Press, p. 13-40.
- Drumm, H. J. (1994) Theoretical and ethical foundations of human resource management: a German point of view. *Employee Relations*, 16 (4), p. pp.35-47.
- Dunning, J. H. and Lundan, S. M. (2008) *Multinational enterprises and the global economy*. Cheltenham, Edward Elgar Publishing Ltd.
- Edstroem, A. and Galbraith, J. R. (1977) Transfer of managers as a coordination and control strategy in multinational organizations. *Administrative Science Quarterly*, 22 (2), p. 248-263.
- Edwards, T. (1998) Multinationals, labour management and the process of reverse diffusion: A case study. *The International Journal of Human Resource Management*, 9 (4), p. 696 709.
- Edwards, T., et al. (2007) Conceptual approaches to the transfer of employment practices in multinational companies: an integrated approach. *Human Resource Management Journal*, 17 (3), p. 201-217.
- Edwards, T. and Ferner, A. (2002) The renewed 'American Challenge': a review of employment practice in US multinationals. *Industrial Relations Journal*, 33 (2), p. 94-111.
- Endo, K. (2006) Foreign direct investment in tourism flows and volumes. *Tourism Management*, 27 (4), p. 600-614.
- Feleppa, R. (1986) Emics, etics, and social objectivity. *Current Anthropology*, 27 (3), p. 243-255.
- Ferner, A. (1997) Country of origin effects and HRM in multinational companies. *Human Resource Management Journal*, 7 (1), p. 19-37.
- Ferner, A., et al. (2004) Dynamics of central control and subsidiary autonomy in the management of human resources: case-study evidence from US MNCs in the UK. Organization Studies, 25 (3), p. 363-391.
- Ferner, A., et al. (2005) Institutional theory and the cross-national transfer of employment policy: the case of 'workforce diversity' in US multinationals. *Journal of International Business Studies*, 36 (3), p. 304-321.
- Ferner, A. and Edwards, P. (1995) Power and the diffusion of organizational change within multinational enterprises. *European Journal of Industrial Relations*, 1 (2), p. 229-257.
- Ferner, A., et al. (2001) Country-of-origin effects, host-country effects, and the management of HR in multinationals: German companies in Britain and Spain. *Journal of World Business*, 36 (2), p. 107-127.

- Ferner, A. and Varul, M. Z. (2000) Internationalisation and the personnel function in German multinationals. *Human Resource Management Journal*, 10 (3), p. 79-96.
- Ford, R. C. and Heaton, C. P. (2001) Lessons from hospitality that can serve anyone. *Organizational Dynamics*, 30 (1), p. 30-47.
- Fukuyama, F. (1995) *Trust: social virtues and the creation of prosperity.* New York, Free Press.
- Geppert, M. and Clark, E. (2003) Knowledge and learning in transnational ventures: an actor-centred approach. *Management Decision*, 41 (5), p. 433-442.
- Geringer, J. M., et al. (2002) In search of "best practices" in international human resource management: research design and methodology. *Human Resource Management*, 41 (1), p. 5-30.
- Glover, L. and Wilkinson, A. (2007) Worlds colliding: the translation of modern management practices within a UK based subsidiary of a Korean-owned MNC. *The International Journal of Human Resource Management*, 18 (8), p. 1437 - 1455.
- Gudykunst, W. B. (1997) Cultural variability in communication. *Communication Research*, 24 (4), p. 327-348.
- Guillet, B. D., et al. (2011) Interpreting the mind of multinational hotel investors: Future trends and implications in China. *International Journal of Hospitality Management*, 30 (2), p. 222-232.
- Harris, S. G. (1994) Organizational culture and individual sensemaking: A schemabased perspective. *Organization Science*, 5 (3), p. 309-321.
- Haveman, H. A. (1993) Follow the leader: mimetic isomorphism and entry into new markets. *Administrative Science Quarterly*, 38 (4), p. 593-627.
- Haynes, P. and Fryer, G. (2000) Human resources, service quality and performance: a case study. *International Journal of Contemporary Hospitality Management*, 12 (4), p. 240-248.
- Hejazi, W. (2007) Reconsidering the concentration of US MNE activity: Is it global, regional or national? *Management International Review*, 47 (1), p. 5-27.
- Hjalager, A.-M. (2007) Stages in the economic globalization of tourism. *Annals of Tourism Research*, 34 (2), p. 437-457.
- Hofstede, G. (1980) Motivation, leadership, and organization: do American theories apply abroad? *Organizational Dynamics*, 9 (1), p. 42-63.
- Hofstede, G. (1984) *Culture's consequences: international differences in work-related values.* Newbury Park, Sage.
- Hofstede, G. (1993) Cultural constraints in management theories. Academy of Management Executive, 7 (1), p. 81-94.
- Hofstede, G. (2002) Dimensions do not exist: a reply to Brendan McSweeney. *Human Relations*, 55 (11), p. 1355-1361.
- Hofstede, G. and McCrae, R. R. (2004) Personality and culture revisited: linking traits and dimensions of culture. *Cross-Cultural Research*, 38 (1), p. 52-88.
- Jackson, G. (2010) Actors and institutions. *In*: Morgan, G., et al. (eds.) *The Oxford handbook of comparative institutional analysis*. Oxford, Oxford University Press, p. 63-86.

- Jackson, S. E. and Schuler, R. S. (1995) Understanding human resource management in the context of organizations and their environments. *Annual Review of Psychology*, 46 (1), p. 237-264.
- Jindra, B., et al. (2009) Subsidiary roles, vertical linkages and economic development: lessons from transition economies. *Journal of World Business*, 44 (2), p. 167-179.
- Johnson, C. and Vanetti, M. (2005) Locational strategies of international hotel chains. Annals of Tourism Research, 32 (4), p. 1077-1099.
- Kelliher, C. and Johnson, K. (1997) Personnel management in hotels an update: a move to human resource management. *Progress in Tourism and Hospitality Research*, 3 p. 321-331.
- Knox, A. (2010) 'Lost in translation': an analysis of temporary work agency employment in hotels. *Work, Employment & Society*, 24 (3), p. 449-467.
- Kostova, T. (1999) Transnational transfer of strategic organizational practices: a contextual perspective. *The Academy of Management Review*, 24 (2), p. 308-324.
- Kostova, T. and Roth, K. (2002) Adoption of an organizational practice by subsidiaries of multinational corporations: institutional and relational effects. *The Academy of Management Journal*, 45 (1), p. 215-233.
- Kundu, S. K. and Contractor, F. J. (1999) Country location choices of service multinationals: An empirical study of the international hotel sector. *Journal* of International Management, 5 (4), p. 299-317.
- Lane, C. (2006) Institutional transformation and system change: changes in the corporate governance of German corporations. *In*: Morgan, G., et al. (eds.) *Changing Capitalisms? Internationalism, institutional change, and systems of economic organization*. p.
- Lashley, C. (1998) Matching the management of human resources to service operations. . *International Journal of Contemporary Hospitality Management*, 10 (24), p. 24-33.
- Lau, C.-M. and Ngo, H.-Y. (1996) One country many cultures: organizational cultures of firms of different country origins. *International Business Review*, 5 (5), p. 469-486.
- Lewis, C. (2008) *How the East was won: the impact of multinational companies in Eastern Europe and the former Soviet Union (1989-2004).* London, Coptic.
- Littlejohn, D. (1997) Internationalization in hotels: current aspects and developments. International Journal of Contemporary Hospitality Management, 9 (5/6), p. 187–198.
- Losekoot, E., et al. (2001) Conceptualising and operationalising the research interface between facilities management and hospitality management. *Facilities*, 19 (7/8), p. 296-303.
- Lucas, R. and Deery, M. (2004) Significant developments and emerging issues in human resource management. *International Journal of Hospitality Management*, 23 (5), p. 459-472.
- Lukka, K. and Modell, S. (2010) Validation in interpretive management accounting research. *Accounting, Organizations and Society*, 35 (4), p. 462-477.
- Maher, A. (1993) Measuring human resource value: an analysis based on the hotel industry. *Journal of Human Resource Costing & Accounting*, 1 (1), p. 15-33.

- Maheswaran, D. and Sharon, S. (2000) Issues and new directions in global consumer psychology. *Journal of Consumer Psychology*, 9 (2), p. 59-66.
- Martin, G. and Beaumont, P. (1998) Diffusing 'best practice' in multinational firms: prospects, practice and contestation. *The International Journal of Human Resource Management*, 9 (4), p. 671 - 695.
- Mathews, V. E. (2000) Competition in the international hotel industry. *International Journal of Contemporary Hospitality Management*, 12 (2), p. 114-118.
- McDowell, L., et al. (2007) Division, segmentation, and interpellation: the embodied labors of migrant workers in a Greater London hotel. *Economic Geography*, 83 (1), p. 1-25.
- McEvoy, G. (1984) Small business personnel practices. *Journal of Small Business* Management, 23 (4), p. 1-8.
- Minbaeva, D., et al. (2003) MNC knowledge transfer, subsidiary absorptive capacity, and HRM. *Journal of International Business Studies*, 34 (6), p. 586-599.
- Morris, M. W., et al. (1999) Views from inside and outside: integrating emic and etic insights about culture and justice judgment. *The Academy of Management Review*, 24 (4), p. 781-796
- Muller, M. (1999) Human Resource Management under institutional constraints: the case of Germany. *British Journal of Management*, 10, p. S31-S44.
- Myloni, B., et al. (2007) The effect of corporate-level organizational factors on the transfer of human resource management practices: European and US MNCs and their Greek subsidiaries. *International Journal of Human Resource Management*, 18 (12), p. 2057-2074.
- Nankervis, A. R. (2000) Human Resource Management strategies as competitive advantage: a case example of the hospitality sector in Southeast Asia & the Pacific Rim. *Research and Practice in Human Resource Management*, 8 (1), p. 111-133.
- Nickson, D. (1997) Research: "colorful stories" or historical insight?-A review of the auto/biographies of Charles Forte, Conrad Hilton, J.W. Marriott and Kemmons Wilson. *Journal of Hospitality & Tourism Research*, 21 (1), p. 179-192.
- Ohmae, K. (1990) The borderless world. New York, Harper Collins.
- Ohmae, K. (1996) The end of the nation state. Cambridge MA, Free Press.
- Pender, L. (2005) Managing the tourism system. *In*: Pender, L. and Sharpley, R. (eds.) *The management of tourism.* London, Sage, p. 1-13.
- Perlmutter, H. V. (1969) The tortuous evolution of the multinational corporation. *Columbia Journal of World Business*, 4 (1), p. 9-18.
- Peterson, M. F. and Pike, K. L. (2002) Emics and etics for organizational studies. International Journal of Cross Cultural Management, 2 (1), p. 5-19.
- Pine, R., et al. (2000) The challenges and opportunities of franchising in China's hotel industry. *International Journal of Contemporary Hospitality Management*, 12 (5), p. 300-307.
- Pudelko, M. and Harzing, A.-W. (2007) Country-of-origin, localization, or dominance effect? An empirical investigation of HRM practices in foreign subsidiaries. *Human Resource Management*, 46 (4), p. 535-559.

- Pullman, M. E., et al. (2001) Service design and operations strategy formulation in multicultural markets. *Journal of Operations Management*, 19 (2), p. 239-254.
- Redman, T. and Mathews, B. P. (1998) Service quality and human resource management: a review and research agenda. *Personnel Review*, 27 (1), p. 57-77.
- Reid, K., et al. (2005) Exploring lived experience. The Psychologist, 18 (1), p. 20-23.
- Reisinger, Y. (2001) Unique characteristics of tourism, hospitality, and leisure services. *In*: Kandampully, J., et al. (eds.) *Service quality management in hospitality, tourism, and leisure*. Binghamton, Haworth Press, p. 15-50.
- Ritzer, G. (1998) The McDonaldization thesis: explorations and extensions. London, Sage.
- Roper, A., et al. (1997) The multi-cultural management of international hotel groups. *International Journal of Hospitality Management*, 16 (2), p. 147-159.
- Rosenzweig, P. (2006) The dual logics behind international human resource management: pressures for global integration and local responsiveness. *In:* Björkman, I. and Stahl, G. (eds.) *Handbook of Research in International Human Resource Management.* Cheltenham, Edward Elgar, p.
- Rosenzweig, P. M. and Nohria, N. (1994) Influences on human resource management practices in multinational corporations. *Journal of International Business Studies*, 25 (2), p. 229-251.
- Rosenzweig, P. M. and Singh, J. V. (1991) Organizational environments and the multinational enterprise. *The Academy of Management Review*, 16 (2), p. 340-361.
- Rowley, G. and Purcell, K. (2001) 'As cooks go, she went': is labour churn inevitable? *International Journal of Hospitality Management*, 20 (2), p. 163-185.
- Royle, T. and Ortiz, L. (2009) Dominance effects from local competitors: setting institutional parameters for employment relations in multinational subsidiaries; a case from the Spanish supermarket sector. *British Journal of Industrial Relations*, 47 (4), p. 653-675.
- Rutherford, D. G. and O'Fallon, M. J. (2007) *Hotel management and operations*. Hoboken, NJ, John Wiley & Sons.
- Saunders, M., et al. (2007) *Research methods for business students*. Harlow, Pearson Education UK.
- Schnitzer, M. (1999) Expropriation and control rights: A dynamic model of foreign direct investment. *International Journal of Industrial Organization*, 17 (8), p. 1113-1137.
- Sisson, K., et al. (2003) All benchmarkers now? Benchmarking and the 'Europeanisation' of industrial relations. *Industrial Relations Journal*, 34 (1), p. 15-31.
- Szulanski, G. (1996) Exploring internal stickiness: impediments to the transfer of best practice within the firm. *Strategic Management Journal*, 17, p. 27-43.
- Taylor, S., et al. (1996) Toward an integrative model of strategic international human resource management. *The Academy of Management Review*, 21 (4), p. 959-985.
- Teare, R. (1993) Designing a contemporary hotel service culture. *International Journal of Service Industry Management*, 4 (2), p. 63–73.

- Thite, M., et al. (2011) Internationalization and HRM strategies across subsidiaries in multinational corporations from emerging economies: a conceptual framework. *Journal of World Business*, In Press, Corrected Proof.
- Timo, N. and Davidson, M. (2005) A survey of employee relations practices and demographics of MNC chain and domestic luxury hotels in Australia. *Employee Relations*, 27 (2), p. 175-192.
- Tuunanen, M., et al. (2011) A model of optimal international market expansion: the case of US hotel chains expansion into China. *In: New Developments in the Theory of Networks*. Physica-Verlag HD, p. 135-154.
- Whitla, P., et al. (2007) Global strategies in the international hotel industry. International Journal of Hospitality Management, 26 (4), p. 777-792.
- Whitley, R. (1999) *Divergent capitalisms: the social structuring and change of business systems*. Oxford, Oxford University Press.
- Whitley, R. (2006) Understanding differences: searching for the social processes that construct and reproduce variety in science and economic organization. *Organization Studies*, 27 (8), p. 1153-1177.
- Wright, P. M., et al. (2005) New models of strategic HRM in a global context. *The International Journal of Human Resource Management*, 16 (6), p. 875 - 881.
- WTO (2001) *Tourism 2020 Vision: global forecasts and profiles of market segments.* Madrid, WTO.
- WTO (2011) International tourism 2010: multi-speed recovery. [online]. Available from: http://85.62.13.114/media/news/en/press_det.php?id=7331&idioma=E [Accessed 27/06/2011].
- WTTC (2009) Travel & tourism economic impact: executive summary. [online]. Available from: http://www.wttc.org/bin/pdf/original_pdf_file/exec_summary_2009.pdf [Accessed 07/07/2010].
- WTTC (2011) *World -key facts at a glance*. [online]. Available from: http://www.wttc.org/eng/Tourism_Research/Economic_Research/ [Accessed 17/02/2011].

Understanding Training in a Complexity Theory Context: Training Process as a Simplifier of Organizational Complexity

Cristina – Catalina Chirita¹, Alexandros Psychogios²

 ¹ Business Administration and Economics Department, CITY College, International Faculty of the University of Sheffield, cristina.chirita@city.academic.gr
 ² Business Administration and Economics Department, CITY College, International Faculty

² Business Administration and Economics Department, CITY College, International Faculty of the University of Sheffield, a.psychogios@city.academic.gr

Abstract. The purpose of this paper is to examine training, as a learning process and knowledge acquiring activity, from a complexity theoretical perspective. In this regard, a qualitative research was conducted, entailing personal interviews with middle level managers from subsidiaries of multinational companies (MNCs) in Romania, aiming to explore managers' attitudes towards the impact of training as a practice that can influence employees' patterns of behavior in coping with current uncertainty. Findings revealed that managers seem to agree with the recent approach of complexity theory as being applicable at organizational level. Moreover, training is perceived as an important practice aiming at increasing employee involvement and commitment towards organization. Finally, training is perceived as a practice that, up to an extent, aims to emphasize patterns of behavior that enable employees to adapt to the turbulent environment. Therefore, this paper is expected to bring a contribution in extending the current understanding of learning process in organizations, by emphasizing that the more managers can understand the latter scope of training the more their organizations can deal with aspects of complexity.

Keywords: *complexity, managers, multinational companies, training.*

1. Introduction

Currently, global economy faces fundamental changes, caused by an increase of actors interacting within different industries, constantly changing information and depending on each other, which determine business environment to have an increasing complex character and force managers to take fast and appropriate decisions (Nedopil et al., 2011). The context of the research is represented by the current character of businesses, which are considered complex systems attempting to adapt to the accelerated changing environment (Teisman and Klijn, 2008).

The need for this study emerged by the necessity of understanding the business reality and exploring the pathways managers follow when dealing with complexity. In an effort of revealing solutions able to assist them in coping with turbulent times, managers engage into learning activities, since knowledge acquiring is considered the main processes underlying human evolution, and training is a common practice seeking to engage employees in learning experiences within working environment.

Therefore, this paper aims to identify whether training can be perceived by managers as a simplifier of organizational complexity. In this respect, after analyzing the literature on this topic, the ideas that came into sight rose to **the research objectives of this paper**, presented as follows:

- a. To identify insights regarding the managerial perspective of complexity theory in coping with unpredictable times.
- b. To explore how training process has an impact on employees' desirable attitudes perceived by managers to improve organizational performance.
- c. To determine outcomes of employees training in regard to patterns of behavior necessary for adapting to the turbulent environment.

The structure of this paper contains four sections. The first section examines briefly but comprehensively the literature attempting to offer an insight of the topics of complexity theory and training process, summarizing the limitations and gaps existent. The second section describes the methodology followed, while the third one analyses the results obtained. The final section discusses the key findings by analyzing similarities and contradictions between literature and research results.

2. Literature Review

2.1. Understanding complexity theory

Nowadays, the increasing number of actors playing within the global economy, interacting and depending on each other, determined a so-called flat-world (Nedopil et al., 2011), where information are ambiguous, cause-effect relationships are reciprocal and leaders are in difficulty when comes to take fast proper decisions. In an attempt to study the dynamic phenomena that businesses face, researchers transposed the main characteristics of living systems to business organizations (Teisman and Klijn, 2008), which are composed of different categories of actors and processes, whose behaviors, attitudes and flows can be observed under the influence of external

factors or internal drivers. All this makes the world increasingly complex (Draman, 2004; Nedopil et al., 2011), and the questions that arise refer to what the concept of complexity really means and how managers can cope with or even overcome complexity.

Complexity is defined as a study (Devies et al., 2007), a methodology or a philosophy of science that sees phenomena and events around connected one to another (Goulielmos, 2002), that cannot be explained by traditional scientific methods (Hayes, 2008), or a way of transforming parts of a whole from complicated and disordered links into simple and easy to understand ones (Goulielmos, 2002), and, conversely, a way of creating complex structures from simply underlying rules (Devies et al., 2007), where independent agents can spontaneously organize themselves into a coherent system.

However, the concept of complexity itself is difficult and complex to be defined and described, and researchers did not arrive yet to a unanimous agreement regarding its meaning, so there is not any universally accepted definition of the term (Nedopil et al., 2011). But what should be mentioned though, is that several characteristics of the concept were agreed upon by many researchers.

Nedopil et al. (2011) argued that complexity is characterized by a great number of component parts with a significant interaction between them, by nonlinearity and broken symmetry, whereas for Teisman (2008), context or landscape, self-organization and nonlinear dynamics are the main attributes of this concept. The context is important since it changes continually, self-organization is the capacity of component parts to auto-arrange between them, and nonlinear evolution implies an unpredictable development.

Therefore, the concept of **complex adaptive systems** is adopted by many practitioners (Bloch, 2005), referring to entities that consist of a large number of independent agents, moderately connected, which can self-organize themselves into coherent systems (Goulielmos, 2002; Mischen and Jackson, 2008), when simple rules are applied (Mason, 2008). Studies showed that complex adaptive systems have some certain common characteristics: they consist of a number of inter-linked and internetworked self-organized and heterogeneous components and they create a great diversity (Bloch, 2005; Devies et al., 2007); each of the components behaves dynamically and nonlinear on its own rules (Huang, 2009; Teisman, 2008); they are largely unpredictable and have a sensitive dependence, so small changes in initial conditions may result in big differences in effects (Huang, 2009; Peterson and Meckler, 2001); they search for fitness peaks and landscapes (Bloch, 2005; Mischen and Jackson 2008).

Considering all these characteristics and referring to the theory previously presented that approaches businesses as systems, and, implicitly, as complex adaptive systems, it can be argued that to human organizations all these attributes can be applicable. Therefore, a human organization is composed from self-organized elements consisting of employees, as individuals organized in departments and divisions, which under changeable circumstances can behave unpredictable, tending to adapt to the landscape.

However, when the process of adaptation is not possible or it is difficult to be realized on short term, the process of co-evolvement in the environment is preferred, therefore, human organizations can be considered also **complex evolving systems**. So the purpose of research studies conducted on this topic was to find whether there can be ways to efficiently manage these issues faced by organizations (Nedopil et al., 2011). Considering businesses as complex adaptive systems, from the managerial perspective it can be argued that in order to adapt to current turbulent and changing environments, unpredictable behaviors and structures might emerge from simple underlying rules (Devies et al., 2007), which might not be centrally and effectively controlled by managers (Mason, 2008).

In this respect, a subset of complexity theory emerged, which stated that small changes in initial conditions of a system can create large changes at a later date (Hayes, 2007). The **chaos** theory, also known as "the butterfly effect" – if a butterfly flaps its wings in a place on earth might cause a storm in another place on earth (Dolan et al., 2003) – was adopted as an unpredictable behavior, with great sensitivity at small changes of the initial conditions, or even a complete disorder or confusion (Burnes, 2004; Huang, 2009; Obolensky, 2008).

There is an entire debate in regard to defining order and randomness, where Burnes (2004) stated that the randomness of chaos, from the complexity perspective, can be seen as a different way of order, Huang (2009) argued that the hidden order is characterized by strange attractors, whilst Goulielmos (2002) stated that complexity considers the organized activity as a chaotic order, being composed from chaos and order in the same time. Therefore, chaos is a system characterized by nonlinear and dynamic relationships (Burnes, 2004; Huang, 2009), and it shows that the interaction of components of an even fairly simple system can create an unexpected result, in the circumstances of certain conditions (Hayes, 2007; Resnicow and Page, 2008), where patterns of behavior may unfold in irregular but similar forms (Dolan et al., 2003).

A proper way of managing organizations by complexity and chaos theory suggests that flexibility and adaptation are essential (Levy, 1994), since long-term planning is mostly not possible due to unexpected changes that may occur, but short-term planning can be undertaken (Dimitrov, 2009; Gheorghita, 2010), considering also the degree order that chaotic systems manifest (Dolan et al., 2003).

Nedopil et al. (2011) provided even some guidelines regarding simplifying the organizational complexity to a manageable level, called **complexity simplifiers**, aiming to help managers reach their targets in current turbulent times. These key issues refer to: **alignment of common behaviors** unitarily applicable to the entire organization in turbulent times; **focus**, meaning that the attention should be paid on the aspects that are most important for the organization; **decentralization** refers to delegating authority of decision-making to lower level managers; **standardization of core processes** occurs within certain departments, when transparency and accountability is needed, together with a common language in order to avoid confusions and misunderstandings generated by the turbulent environment. However, there is not much research to probe the application of these simplifiers as being effective in coping with complexity, therefore an important literature gap can be exploited from this aspect.

Managing by complexity theory is a very delicate process, where everything depends on everything else and all are responsible for achieving common goals (Goulielmos, 2002), and it is argued to be concentrated in coordinating people to developing their skills for personal initiative, self-organization, self-action and

continuous change, in order for the managers to cope with increased uncertainty (Mason, 2008).

Since business organizations are seen as living systems, relationships and communication are now, more than ever, a key factor in enabling and enhancing progress, being even more important than structures and processes (Goulielmos, 2002). The questions arisen at this point refer to explaining to which extent managers are able to adapt to current complexity and by which means they can enhance these necessary core competencies.

2.2. Training process

Training represents an investment in human capital that businesses make in order to enhance their employees' skills and abilities, increase their productivity and overall improve the organizational performance (Landeta et al., 2009), so it is expected that the investment to be recovered later. Due to fast technological developments and economical changes, investment in employee training is a factor of growth, being also an essential element of competition, although in the current economic crisis, human resources are the first affected, by reducing the budget for education and training.

It was showed that there are some certain pathways through which training contributes to the overall learning experiences results, and these are reflected both at the individual level and at the inter-individual level. As regards individual level, training programs contribute to developing application knowledge and business context knowledge. Regarding the inter-individual level, transactive memory is enhanced and collaborative task knowledge is improved (Sharma and Yetton, 2007).

In this regard, researchers showed that organizations providing training programs noticed that their employees had more positive attitudes towards the company and managers (Landeta et al., 2009), which had also a favorable influence on employee job satisfaction, commitment and overall performance.

2.3. The rationale and the context of the study

Considering that complex adaptive systems have a sensitive dependency to small changes (Huang, 2009), it can be argued that employees, which are the main internetworked and heterogeneous components, might react differently and unpredictable when aligning them to the same rules and procedures, and they might tend to self-organize and to develop some patterns of behavior in order to adapt to these rules. Furthermore, training was analyzed as a tool of reinforcing skills (Sharma and Yetton, 2007) and reminding, from time to time, the rules and conditions under which the activity of the company is conducted. The question arisen now is whether training is approached by the managers as such tool, and to which extent it can help in generating or enhancing certain desirable patterns of behavior to employees, in terms of following or breaking the rules from their routine, so whether training can be seen as a simplifier of organizational complexity and to which extent.

Therefore, training process, from the perspective of complexity management, is still questioned in regard to the outcomes provided, and there is limited research that can show whether it influences or not the trainees patterns of behavior in terms of following or breaking the rules of their order, since they are considered independent agents but moderately connected to the same chaotic and complex adaptive systems, so whether training can be seen as a simplifier of organizational performance.

The context of the study is represented by MNCs operating in a rather complex and turbulent business context like the one of Romania. There is less literature concerning the application of a management practices in such environments, especially from the complexity theory point of view. Romania emerged for 22 years from a long period of centralized economy of around 45 years, having still a small understanding of management under free market conditions and capitalist economy, being still considered a developing country (Radaceanu, 2004). However, with the occasion of the accession into the European Union in 2007, an abundance of international companies selected Romania for opening new subsidiaries (Popa, 2008).

At the end of 2010, there were identified more than 15,200 active groups of subsidiaries of MNCs, representing around 76% of all the businesses in Romania (I.N.S., 2011). The most preferred business sectors for MNCs are commerce (retail), technical (manufacturing and services), real estate, IT, telecommunications, and finance (banking and insurances).

Furthermore, it was argued that the **managerial activity in subsidiaries of MNCs** depends mainly on two aspects (Russu, 2010). The first one regards the **managerial practices** adopted for decision-making processes, and the other one refers to the **knowledge** applied within a company, based on specialty training.

Unfortunately, there is not much empirical evidence within the literature, connecting these two aspects in effectively managing subsidiaries of MNCs in Romania. Therefore, this study seeks to approach training within MNCs operating in the turbulent business environment from the complexity theory perspective. The overall aim is to understand **training as a learning generator and complexity simplifier.**

3. Data and Methodology

The overall **research scope** of this study is to explore whether training is perceived by managers as a simplifier of organizational complexity, and in particular this aspect can be distinguished in five **research questions**:

- a. How do managers approach complexity theory in coping with unpredictable times?
- b. To which extent do managers consider different solutions or simplifiers as applicable in effectively coping with complexity?
- c. To which extent has training an impact on employees' desirable attitudes perceived at the managerial level to improve organizational performance?
- d. Which are the outcomes of employee training in regard to patterns of behavior necessary for adapting to the turbulent environment?
- e. To which extent has training a role on repeating or changing the rules of the complex system employees are part of, and how do they apply their learning experiences in this regard?

Due to the complex and sensitive nature of the topic revealing social phenomena, a **qualitative research strategy** was preferred, being the proper way to investigate the particular research topic that cannot be easily quantified. Non-standardized, semi-

structured, face-to-face **interviews** were conducted, argued to be the most proper way of seeking for new insights through open discussion with people having managerial positions, and to have the advantage of getting much insight into the phenomena.

Moreover, a **single qualitative data collection technique** was used and the population targeted was formed by middle level managers from subsidiaries of MNCs located in Romania, activating in four different business sectors (commerce, finance, technical and telecommunications), because this type of businesses counted around three-quarters of the Romanian businesses at the end of 2010 (I.N.S., 2011), and this category of managers are more related with the process of training for employees.

Additionally, the **snowballing sampling technique** was preferred, since it was difficult to identify representatives of the desired population (Collins and Hussey, 2003; Saunders et al., 2009). However, the criteria of selecting these participants was to lead employees that have participated in training programs in the last one-two years and to have various attributes in terms of age and work experience.

After collecting the data, the analysis was accomplished by summarizing and categorizing information gathered into groups of meanings and patterns through the use of conceptualization. However, there was the barrier of language, as the interviews were held in Romanian, whilst the whole study was designed and analyzed in English.

4. Empirical Analysis

Demographic characteristics of interviewees are shown in **Appendix 1**. Due to various business sectors managers interviewed come from, there were observed some differences in their answers, but similarities around those belonging to the same industry. Therefore, the findings are concentrated around each industry, although it is not the case of generalizing findings at the level of business sector. The findings are concentrated in three main areas, referring to managers' attitudes towards complexity theory, training as a mean of developing performance-oriented behaviors, and training as a practice to emphasize patterns of behavior for adapting to turbulent environments.

4.1. Managers' attitudes towards complexity theory

When questioned about the general approach towards organization as being a complex adaptive system, it was unanimously answered that businesses can be easily compared with any other living system, characterized by typical structures and components, by unpredictable behaviors of employees that might not be unitarily and effectively controlled by managers in crisis situations, and by specific procedures and norms underlying the whole activity and functioning of the organization.

"Plenty of rules and norms we have to align to, frequent turbulences on the financial market that affect our daily activities, decision we have to take fast, all these I believe makes everything to be complex here." (I. 6 – Branch Manager – Financial sector)

Furthermore, chaos theory is also agreed as a reality underpinning business environment, and managers believe that it is about the small steps of change occurring within a company that make the big differences during time. Also, chaos is perceived mostly as a disorder faced by the organization or within a departmental level, which can generate unpredictability or unexpectedness towards future, so a relationship of causality is seen between the two, and not a synonymy.

When questioned about providing some solutions aiming to help them reduce organizational complexity to a better manageable level, some managers answered that it is important for a company to have core values and common rules underlying its activity, to pay attention to the most crucial details for the organization and stakeholders interests, and to delegate authority of decision-making by involving employees in coping with unpredictable situations.

Most of the questioned managers argued that besides a managerial approach of coping with uncertainty, there is the need of engagement from employees when it comes to crisis situations, and they are preferred to possess some certain abilities and skills needed for a fruitful cooperation in dealing with unstable issues. Managers are not willing and do not expect to face turbulent situations by themselves, therefore they seek for specific skills in employees even from the recruiting process, in order to be sure that when needed, they are prepared to put their capacities into practice.

"When we hire a new employee, we seek for some certain abilities in him or her, because in our sector changes occur often, so we need people to be able to face them effectively and to take proper decisions when needed. [..] But we realize that these abilities can be enhanced also." (I. 2 – General Manager – Technical sector)

Nevertheless, managers argued also that desirable skills can be though improved by learning, so through employee participation into training programs suitable designed and appropriately held.

4.2. Training as a mean of developing performance-oriented behaviors

Firstly, it should be mentioned that when asked whether they consider training as a cost or an investment for the company, managers responded in unanimity that it is an investment aiming to give in return benefits for both individuals and organization, on short or long term, depending on the results pursued.

Furthermore, training outcomes were argued to be evaluated by managers through open discussions with employees providing feedback regarding their learning experiences, and, on a longer period, through observing employees improvements in achieving their working tasks, and consequently, in enhancing employee performance.

Two main directions emerged from this point, and they are related to explaining how performance is perceived by managers and to identifying factors that determine employee performance. In regard to perceiving performance, it was stated that as long as individual and organizational goals are achieved, an employee is considered performing, however, there are also taken into account subjective aspects of its relationship with the whole organization and the processes undertaken in reaching the proposed objectives. In regard to determining factors for achieving performance, two main types were identified, referring to commitment and competence.

"Since we have a very bureaucratic structure, for us performance is simple: the amount of clients attracted and their spending for insurances." (I. 5 - Risk Manager - Financial sector)

As regards the role of training to create a more committed workforce and increased trust within the organization, managers unanimously responded they perceive employees to display more engagement, although open discussions are usually not held in this regard, expressing the level of motivation being a sensitive matter.

Concerning the role of training in providing new skills or improving existing ones, necessary to do the job better, different opinions were gathered, due to particular characteristics of business sectors. Various skills were mentioned as being important in enhancing performance, referring to: personal initiative, self-organizing ability, responsibility, results orientation, objective self-assessment, perseverance, willing for continual development, thinking outside the box, empathy and capacity for continuous adapting to change.

However, it is debated that these important skills are hardly to be developed only through training programs, due to a high level of subjectivity and relativity accompanying these abilities, so there must be also additional learning processes in support, like coaching for instance, and also due to employees' particular styles and capacities of learning and processing new information acquired.

"You can expect from people to come up with solutions, to have personal initiative, to take proper decisions by themselves, but only training cannot teach them all these. In our sector, it is important to have your manager beside you, as a mentor let's say, and to "steal" practice from others." (I. 15 - HR Manager – Commerce)

Summarizing, it can be argued that training plays an important role in both enhancing employee involvement and commitment within the organizational critical situations, and improving desirable abilities of doing better in the job to an extent. Considering these two perspectives as main factors in determining employee performance, it can be concluded though that training plays an important role in this regard, and consequently, in enhancing overall organizational performance.

4.3. Training as a practice to emphasize patterns of behavior for adapting to turbulent environments

Concerning the pathways through which training enhanced learning experiences results and whether these were reflected to the individual and inter-individual level after the completion of the training program, most of the managers agreed that the trainer plays an important role in influencing the outcomes of the training program. Therefore, the purpose of the trainer is to align trainees, as much as possible, to common rules and conditions, which emphasized enough, may tend to have a better influence on common and desirable learning outcomes.

On another hand, when assessing learning results, besides individual experiences gathered that have an impact on the future attitude towards work and organization, an accent should be also put on the inter-individual level. Managers stated that an important outcome after a training process is in regard to sharing experiences with peers, and that communication is essential within a training practice, and its results are significantly influenced by the way communication process is guided.

"After the previous training, my employees said that they finally found the solution of an issue, however not due to participating in the program, but because of meeting some colleagues from another branch, which shared from their common experience." (I. 5 - Risk Manager - Technical sector)

Approaching the aspect of rules and procedures underlying the functioning of an effective system, opinions were divided. In finance and technical sector, due to bureaucratic norms, managers stated that at a surface level, as many directives are provided, as more clearly the activity seems like, and as less disorder and ambiguity are expected. However, practice showed that unpredictable situations appeared regardless efforts of managers to avoid unexpectedness. As regards the necessary ingredients in reducing complexity, managers answered that an apparent and brittle stability lies actually on uncertainty, ready to reveal in any moment, drawn even by a drop of change. It can be explained, thus, that simple underlying rules tend to increase certainty and reduce complexity, but when procedures tend to become abundant and suffocating, the effect is contrary, so the complexity is even more enhanced.

Furthermore, responses on whether participating in such programs has an influence on employees' patterns of behavior towards rules from their routine were various, but can be grouped in three main directions. In commerce, managers argued that employees returning from training programs tend to break rules already known, due to regular changes occurred within their sector, so to divert from their daily routine in order to adapt to the turbulent environment and evolve with it. In technical sector and finance, it was stated that the role of trainings within their areas is to remind, from time to time, the general rules that determine the system to function, so an outcome of following these rules is desirable, in order to cope with the changing environment. The third direction was drawn by managers from telecommunications, stating that neither following nor breaking the rules is applicable in their cases, but adjusting the rules, which was explained as a process of identifying the underlying procedures to be followed, and adjusting them to the different organizational considerations.

"In theory it's easy to be advised to follow some universal laws and be effective. But in practice, what we actually do is to adjust the rules to our particularities. There are some common frames to be followed, that's for sure, but it depends on us how we use them to be better for us." (I. 11 - PR Team Leader – Telecommunications sector)

Furthermore, the patterns of behavior identified were argued to be actually part of the outcomes of training, since the general purpose of this practice is to enable employees to have a better understanding of the environment and, implicitly, of the uncertainty, which, if effectively coped with, leads to performance. Thus, it can be argued that training has a role in adapting employees' patterns of behavior to the reality surrounding, and when this aspect is achieved, it leads to a better understanding of how complexity can be better managed and organizational performance improved.

5. Conclusions

Complexity and chaos are theories emerged from the necessity of having a deeper understanding of reality (Nedopil et al., 2011), and their main common characteristic is that large differences in results due to unpredictable radical changes may come from small differences in initial conditions (Goulielmos, 2002; Mischen and Jackson, 2008). Organizations, as complex adaptive systems, need acknowledge their unpredictable nature and that they cannot be controlled (Mason, 2008; Teisman, 2008) and to understand the drivers of complexity (Nedopil et al., 2011).

Therefore, in regard to the **first objective of the study**, **similarities** were observed between the literature and the findings, since managers mentioned several characteristics of complex adaptive systems as being applicable to their organizations. Also, they identified from their experiences factors that may lead to an increased complexity, which strengthen the idea that there are complexity drivers enhancing even more the unpredictable character of the environment (Nedopil et al., 2011), aspects which if known and avoided by managers may enable them reduce uncertainty (Resnicow and Page, 2008). Moreover, some managers provided some solutions aiming to simplify organizational complexity, related to following common rules and core values at an organizational level, concentrating to most important priorities, and engaging employees possessing desirable abilities in coping with turbulent times.

A **contradiction** occurred though in regard to how chaos is perceived at the managerial level, since it is considered very simplistic to be characterized by a general state of disorder appeared as a consequence of lack of rules and unpredictable or improper managed changes. Therefore, it is a relationship of causality between lack of alignment to common rules, as a generator, and disorder or chaos, as a result, idea that contradicts the literature, considering chaos itself a whole system characterized by nonlinear and dynamic relationships (Burnes, 2004; Huang, 2009).

Furthermore, as regards the **second objective of the study**, **similarities** occurred in considering training as an investment for the company, aiming to provide further benefits, and the two main factors discussed to have an important influence in determining employee performance were organizational involvement and commitment, and improving desirable abilities of performing the job better. The most important finding though relates to the desirable attitudes that managers considered to be enhanced through training (Buchner, 2007; Landeta et al., 2009), and besides those mentioned within the literature (Garev, 2010), there were some additional other skills discovered as necessary, such as: responsibility, objective self-assessment, concern for continual development, empathy and thinking outside the box.

As regards the **third objective of the study**, **similarities** with the literature were found in terms of acknowledging training as a knowledge generator, having also considerable impact on mindsets of employees, considered as independent agents but moderately connected to the same chaotic and complex adaptive systems. It was argued that training, a complex process by itself, has the main purpose to influence desirable learning outcomes of trainees, by aligning them to common rules and conditions, and impact behaviors towards work and organization, at an individual and inter-individual.

All the conditions presented were proved to contribute to increasing organizational performance and better coping with turbulent environment, which lead to a simpler organizational complexity. It can be deducted that training, by meeting all the conditions mentioned below, can lead also to enhancing organizational performance and adapting to the chaotic reality, therefore it can be considered that **training practice can be seen as a simplifier of organizational complexity**:

- It aims at **aligning common behaviors and values unitarily** for the entire organization or department involved within the training process;
- It enables, up to an extent, **repeating rules and procedures** used by managers to control better unpredictable situations and chaos;
- It encourages **focusing to the most important aspects** of the organization and of the environment, useful for achieving common goals;
- It allows **delegating authority of decision-making**, through developing desirable attitudes required by managers in order to grant empowerment;
- It facilitates enhancing relationships and communication, as key factors in coping with unexpected circumstances and achieving progress.

As concerns **practical implications**, the conclusions of this study show that managers should pay more attention to the way they set rules and procedures within organizations, and to the way they organize training programs and select employees for attending them. In terms of **limitations**, there is the issue of a small research sample, with only twenty interviews conducted in four different business sectors, so the results cannot be generalizable to an entire industry level. It was also the barrier of language, and the constraint of reliability, since usually respondents tend to present their realities in a way they consider as being perceived better by people outside organization. In regard to **further research**, it can be recommended that a bigger sample of respondents to be considered, through a multi-method research choice, by including focus groups and observations. Also, besides managers, employees and trainers can be approached, and more criteria can be considered, in terms of varied personal characteristics and more business sectors in selecting the population.

References

Bloch, D.P. (2005) Complexity, Chaos, and Nonlinear Dynamics: A New Perspective on Career Development Theory, *Career Development Quarterly*, 53(3), p. 194-207

Buchner, T. (2007) Performance management theory: A look from the performer's perspective with implications for HRD. *Human Resource Development International*, 10(1), p. 59-73

Burnes, B. (2004) Kurt Lewin and complexity theories: back to the future?. *Journal of Change Management*, 4(4), p. 309-325

Collins, J. and Hussey, R. (2003) Business research. A partial guide for undergraduate and postgraduate students. 2^{nd} edition. Hampshire: Palgrave MacMillan

Devies, T.C.; Cicmil, S.; Crawford, L; Richardson, C. (2007) We're Not In Kansas Anymore, Toto: Mapping The Strange Landscape Of Complexity Theory, And Its Relationship To Project Management. *Project Management Journal*, 38(2), p. 50-61

Dimitrov, N.N. (2009) Understanding complexity and chaos management in Southeast European organizations. Unpublished M.Ed, dissertation, City College – International Faculty of the University of Sheffield

Dolan, S.L.; Garcia, S.; Auerbach, A. (2003) Understanding and Managing Chaos in Organizations, *International Journal of Management*, 20(1), p. 23-36

Draman, R. (2004) Organization/business, management, people and complexity – an approach to their integration. *Human Systems Management*, 23(2), p. 101-110

Garev, S. (2009) Towards exploration of complexity and complex management behavior to organizational success in the turbulent business environments: Evidence from the South Eastern European region. Unpublished M.Ed, dissertation, City College – International Faculty of the University of Sheffield

Gheorghita, V. (2010) *Managing complexity aspects in resizing organizations: Downsize process applied in construction industry.* Unpublished M.Ed, dissertation, City College – International Faculty of the University of Sheffield

Goulielmos, A. (2002) Complexity theory applied to management of shipping companies. *Maritime Policy & Management*, 29(4), p. 375-391

Hayes, P. (2007) Chaos, Complexity and Transformations in Social Care Policy in England. *Public Money & Management*, 27(3), p. 199-206

Hayes, P. (2008) Complexity theory and evaluation in public management. *Public Management Review*, 10(3), p. 401-419

Huang, X. (2009) The Influence Of National Culture, History And Institution On Strategic Management In Chinese Firms: A Complexity Based Perspective. *International Journal of Business Studies*, 17(1), p. 1-18

I.N.S. (2011) Romanian National Institute of Statistics Report. [Available online: http://www.insse.ro/cms/files/statistici/comunicate/com_anuale/grup_intrep2009.pdf]

Landeta, J.; Barrutia, J.; Hoyos, J. (2009) Management turnover expectations: a variable to explain company readiness to engage in continuous management training. *The International Journal of Human Resource Management*, 20(1), p. 164-185

Levy, D. (1994) Chaos Theory And Strategy: Theory, Application And Managerial Implications. *Strategic Management Journal*, 15(1), p. 167-178

Mason, R. (2008) Management actions, attitudes to change and perceptions of the external environment. *Journal of General Management*, 34(1), p. 37-53

Mischen, P.A. and Jackson, S.K. (2008) Connecting The Dots: Applying Complexity Theory, Knowledge Management And Social Network Analysis To Policy Implementation. *Public Administration Quarterly*, 32(3), p. 314-338

Nedopil, C.; Steger., U; Amann, W. (2011) *Managing Complexity in Organizations*. Hampshire: Palgrave MacMillan

Obolensky, N. (2008) Chaos Leadership and Polyarchy – countering leadership stress? *University of Exeter - Centre for Leadership Studies*. [Available online: http://centres.exeter.ac.uk/cls/documents/publications/395.pdf]

Peterson, M.F. and Meckler, M.R. (2001) Cuban–American Entrepreneurs: Chance, Complexity and Chaos. *Organization Studies (Walter de Gruyter GmbH & Co. KG.)*, 22(1), p. 31-57

Popa, A. (2008) Investment in human capital for Romania in context of European Enion integration. *Annals of the University of Oradea, Economic Science Series*, 17(2), p. 374-378

Radaceanu, E. (2004) Some Aspects of Management Training in Romania and American Assistance in its Development. *International Journal of Public Administration*, 27(11&12), p. 893-904

Resnicow, K. and Page, S.E. (2008) Embracing Chaos and Complexity: A Quantum Change for Public Health. *American Journal of Public Health*, 98(8), p. 1382-1389

Russu, C. (2010) Romanian Managerial Training System - Current State and Prospects. *Petroleum - Gas University of Ploiesti Bulletin, Economic Sciences Series,* 62(2), p. 21-30

Saunders, M.; Lewis, P.; Thornhill, A. (2009) *Research methods for business students. 5th edition.* Harlow: Pearson Education

Sharma, R. and Yetton, P. (2007) The contingent effects of training, technical complexity, and task interdependence on successful information systems implementation. *MIS Quarterly*, 31(2), p. 219-238

Teisman, G. (2008) Complexity and management of improvement programmes. *Public Management Review*, 10(3), p. 341-359

Teisman, G. and Klijn, E.H. (2008) Complexity theory and public management: An introduction. *Public Management Review*, 10(3), p. 287-29

Exploring the Relationship between the Learning Organization and Business Performance

Bojana Milić1

¹ Faculty of Technical Sciences, University of Novi Sad, boatim@uns.ac.rs

Abstract. The concept of the learning organization has received substantial attention in the academic literature since the learning processes have been emphasized as a source of competitive advantage. Organizations that implement strategies consistent with the learning organization are considered to achieve improved performance. Although, the linkage between learning and business performance has often been assumed, there is an insufficient number of empirical evidence to support this perspective. The purpose of this study is to explore the relationship between the learning organization concept and business performance in information technology (IT) sector in Serbia. In order to assess this association, employees' responses to the Dimension of Learning Organization Questionnaire (DLOQ) instrument will be used along with perceptual measures of business performance.

Keywords: *learning organization, organizational learning, business performance*

1 Introduction

The present business world is characterized by constant and rapid changes in the global economy which have lead to the emergence of a new organizational form known as *learning organization*. The transformation of organizational systems to reflect "processes of learning, behavior change, and performance improvement" (Slater and Narver, 1995, pp. 63 - 64) has become an organizational imperative as the survival of the organization is at stake. According to the theorists, the capacity to learn faster than competitors may be the only sustainable competitive advantage in volatile environments (De Gus, 1988; Stata, 1989). Learning organizations are designed to enable people and organizations to be adaptive to the environment and responsive to change. Today, it is almost impossible to find an industry that is not in a state of dynamic change requiring organizations to focus on continuous learning

activities. Hence, learning is considered to be a fundamental aspect of competitiveness and it is linked with knowledge acquisition and performance improvement. Jones (2004) emphasizes the importance of organizational learning for performance, defining it as a process by which managers try to increase employees' capabilities in order to better understand and manage the organization and its environment, to accept decisions that increase organizational performance on a continuous basis.

However, although the linkage between learning organization and business performance has often been assumed, there is an insufficient number of empirical evidence to support this perspective. In the literature, more attention has been devoted to defining the learning organization than developing and using methods of measurement. This study aims to reduce this gap by examining the relationship between the learning organization concept and business performance.

2 Literature Review

The concept of learning organization has received great attention in the literature since organizational learning processes have been highlighted as a foundation of competitive advantage (Easterby-Smith & Lyles, 2005; Swieringa & Wierdsma, 1992). Peter Senge (1990) stated that the learning organization is an organization where people continually expand their capacity to create the results they truly desire, where new and expansive patterns of thinking are nurtured, where collective aspiration is set free, and where people are continually learning how to learn together. In addition, one of the significant definitions of this concept was defined by Watkins and Marsick (1993). They stated that the learning organization is one that is characterized by continuous learning for continuous organizational improvement, and by the capacity to transform itself. It is a learning organization in which people are aligned based on a common vision through the continuous dialogue process, generating new knowledge used to create innovative products and services for organizational performance improvement. Marquardt (2002) combines elements of the earlier definitions with additional emphasis on knowledge management and action learning, and states that "a learning organization is a company that learns effectively and collectively and continually transforms itself for better management and use of knowledge; empowers people within and outside the organization to learn as they work; and utilizes technology to maximize both learning and production". In recent times, the term learning organization was defined in the broader point of view as "an organization that is organized to scan for information in its environment, by itself creating information and promoting individuals to transform information into knowledge and coordinate this knowledge between the individuals so that new insight is obtained" (Jensen, 2005).

Although there are differences in perspective what exactly does learning organization mean all views directed to the one that in the long–run business performance is measured by long-term survival and growth of the firm.

There are few empirical studies that have examined the relationship between the learning organization concept and a firm's performance (Narver and Slater 1990; Ellinger et al. 2002; Farrell and Oczkowski 2002; Yang, Watkins, and Marsick 2004;

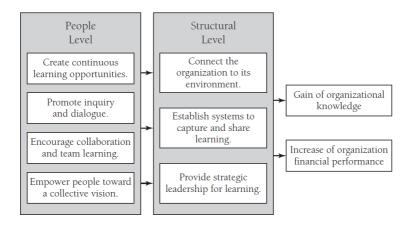
Garcia-Morales, Llorens-Montes, and Verdu-Jover 2006), and have shown significant, positive relationships between these two concepts. The firm's performance included overall performance and a few financial indicators, as well as innovation, new product success, market share and reliable performance.

The primary aim of this study is to assess the relationship between the learning organization concept as articulated by Watkins and Marsick (1993, 1999) and business performance perceived by the employees in IT companies in Serbia.

2.1 Conceptual framework

For the purpose of this paper Watkins and Marsick's (1999) conceptualization of the learning organization was used. They developed the Dimension of Learning Organization Questionnaire (DLOQ) to identify the characteristics of a learning organization (Marsick & Watkins 2003). It is organized into seven sections, as presented in the Figure 1., addressing individual level, team level, and organization level learning.

Figure 1: Conceptual Framework of the Dimension of Learning Organization and Performance Outcomes (Yang, Watkins & Marsick, 2004)



- 1. *Create continuous learning opportunities* learning process is inscribed in the work, so that people can learn in the course of their work (on the job learning). Opportunities for continuous education and growth are provided.
- Promote inquiry and dialogue people gain productive reasoning skills to express their opinions; they have the ability to listen and research into the views of others; the culture of the organization supports questioning, experimentation and provides a feedback.
- 3. *Encourage collaboration and team learning* work is done through the use of employees' groups; members of the groups can become familiar with different ways of thinking, it is expected from the groups to learn and work together; organizational culture supports collaboration and rewards it.

- 4. Create systems to capture and share learning systems for sharing learning are created and used in the work, and employees have free access to accumulated knowledge.
- 5. *Empower people toward a collective vision* people are involved in creating and implementing a shared vision; the responsibility for its implementation rests with those who are involved in decision-making process; people are motivated to learn to truly fulfill their obligations.
- 6. *Connect the organization to its environment* allows people to see the effects of their work from the perspective of the organization; people can use information from environmental organizations and adapt his work; the organization is linked to external communities.
- Provide strategic leadership for learning Leaders shape, improve and support learning; leaders use learning strategically to achieve better business results.

This model was chosen as a framework for this study because of its grounding in practice and its integrative perspective, which emphasizes three key components: (1) systems-level, continuous learning, that (2) is created in order to create and manage knowledge outcomes, which (3) lead to improvement in the organization's performance, and ultimately, its value (Ellinger et al. 2002; Yang, Watkins, and Marsick 2004).

Most of the studies found in the literature that employed the DLOQ instrument to evaluate the seven dimensions of a learning organization and various aspects of business performance have provided generally positive relationship between one or more dimensions of a learning organization and business performance (Ellinger, Ellinger, Yang, & Howton, 2002; Zhang, Zhang & Yang, 2003; Davis & Daley, 2008).

2.2 Research Questions

The study is built on and furthered existing research by examining the relationship between the constructs of the learning organization and business performance. Guiding this study are five research questions:

- 1. To what extent do employees perceive their organizations as a learning organization, employing the Watkins and Marsick DLOQ?
- 2. What is the relationship between the seven dimensions of learning organization and business performance?
- 3. Which dimension of learning organization has the most significant effect on the business performance as perceived by the employees?
- 4. To what degree do the employees' perceptions of their organization as a learning organization differ based on their age, gender, level of employment, and level of education?
- 5. To what degree do the employees' perceptions of business performance differ based on their age, gender, level of employment, and level of education?

By linking the seven dimensions of the learning organization (continuous learning, inquiry and dialogue, collaboration and team learning, empower people, create

systems, connect the organization, and strategic leadership) with the business performance (financial and knowledge performance) it is possible to discover correlations among them that are statistically significant and to provide a new insight on learning organization and performance that has not been considered previously. Besides, results will demonstrate if there is a difference in perception of organization as a learning organization and in perception of business performance, concerning the age, gender, level of employment, and level of education of the respondents.

In addition, this research also affords us the opportunity to further assess the reliability and validity of Watkins and Marsick DLOQ instrument in a different context.

3 Proposed Methodology

A self-administered web-based survey will be used to collect individual-level perception data from employees in a single industry. The use of an employee survey was deemed appropriate to address the proposed research questions.

3.1 Sample

A convenience sample of companies from IT industry in Serbia will be chosen based on accessibility to their top management and their willingness to participate in the survey. IT industry is seen as the most appropriate for this study since it can be considered as representative of knowledge-based industry.

3.2 Instrumentation and procedures

In line with what is described above, author tends to adopt the shorter version of the DLOQ consists of 21 questions to measure the seven learning organization dimensions in the IT sector in Serbia. In addition to these questions, the DLOQ instrument includes two perceptual outcomes measures: perceived financial performance (return on investment, average productivity per employee, time to market, response time for customer complaints, market share and costs per business transaction) and knowledge performance (customer satisfaction, the number of suggestions implemented, the number of new products and services, the percentage of skilled workers compared to the total workforce, the percentage of total spending devoted to technology and information processing, and the number of individuals learning new skills). Respondents will be asked to indicate their perception of the degree to which their organization practices the described behaviors and their assessment of the organization's current performance when compared to the previous year, on a six-point Likert-type scale, where 1 equals the assessment that the behavior 'almost never' occurs, and 6 equals 'almost always'

A translation-back translation procedure will be used to translate the questionnaire from English to Serbian and back to English. Furthermore, pilot testing with a selected group of employees in different companies will be conducted in order to reinforce baseline reliability for the survey before its distribution to the target population. The method of structural equation modeling (SEM) (Benson and Hagtvet, 1996), will be used to examine the demonstrated relationships. Author will use the exploratory sample to search for a fitted model of SEM and test that model in the confirmatory sample. If there is a structural model representing the conceptual framework that fits the data, and the hypothesized relations among the constructs are found to be significant in the desirable directions, then there are significant relations between the proposed seven dimensions of the learning organization and the two organizational outcome variables.

4 Conclusions

Maximizing the learning that occurs in companies is vital for continued success and competitive advantage. The critical issue, however, is being able to identify the most important types and levels of learning within organizations, and in that way to focus scarce time and budgets on organizational learning that has the highest impact on business performance. The findings of this research will have the potential to assist companies who are looking to leverage organizational learning to create higher business performance within their organizations.

The outcomes of this study may also encourage and allow others to conduct research on other factors or elements that may possibly affect the relationship between organizational learning and business performance.

References

Benson, J., & Hagtvet, K. (1996) The interplay among design, data analysis, and theory in the measurement of coping, in Zeidner, M. & Endler, N. (eds), *Handbook of coping*, Wiley, New York, 83–106.

Davis, D. & Daley, B.J. (2008) The learning organization and its dimensions as key factors in firms' performance. *Human Resource Development International*. Vol. 11, no. 1, 51-66, doi:10.1080/13678860701782352.

De Geus, A.P. (1988) Planning as learning. *Harvard Business Review*. Vol. 66, no. 2, 70-4.

Easterby-Smith, M., & Lyles, M. A. (2005) Organizational learning and knowledge management: Agenda for future research. In M. Easterby-Smith & M. A. Lyles (Eds.), *Handbook of organizational learning and knowledge management*, Malden, MA: Blackwell Publishing, 639-652.

Egan, T., Yang, B., & Bartlett, K. R. (2004) The effects of organizational learning culture and job satisfaction on motivation to transfer learning and turnover intention. *Human Resource Development Quarterly*. Vol. 15, no. 3, 279-302.

Ellinger, A. D., Ellinger, A. E., Yang, B., & Howton, S. W. (2002) The relationship between the learning organization concept and firms' financial performance: An empirical assessment. *Human Resource Development Quarterly*, Vol. 13, no. 1, 5-21.

Farrell, M.A., & Oczkowski, E. (2002) Are market orientation and learning orientation necessary for superior organizational performance? *Journal of Market-Focused Management*. Vol. 5, 197–27.

Garcia-Morales, V.J., F.J. Llorens-Montes, & Verdu-Jover, A.J. (2006) Antecedents and consequences of organizational innovation and organizational learning in entrepreneurship. *Industrial Management and Data Systems*. Vol. 106, no. 1, 21–42.

Jensen, P. E. (2005) A contextual theory of learning and the learning organization. *Knowledge and Process Management*. Vol. 12, no. 1, 53-64.

Jones, G. R. (2004) *Organizational theory, design, and change* (4th ed.), NJ: Prentice Hall, Upper Saddle River.

Marquardt, M. (2002) Building the learning organization: mastering the five elements for corporate learning, CA: Davies-Black, Palo Alto.

Narver, J.C., & Slater, S.F. (1990) The effect of a market orientation on business profitability. *Journal of Marketing*. Vol. 54, no. 4, 20–35.

Senge, P.M. (1990) The fifth discipline: the art and practice of the learning organization, Doubleday Currency, New York.

Slater, S.F. & Narver, J.C. (1995) Market orientation and learning organization. *Journal of Marketing.* Vol. 59, 63-74

Stata, R. (1989) Organizational learning: the key to management innovation. *Sloan Management Review*. Vol. 30, no. 3, 63-74.

Swieringa, J., & Wierdsma, A. (1992) *Becoming a learning organization*, Addison-Wesley Publishing Company, Wokingham, England.

Watkins, K. E., & Marsick, V. J. (1993) Sculpting the learning organization: Lessons in the art and science of systemic change, CA: Jossey-Bass, San Francisco. Watkins, K.E., & Marsick, V.J. (1999) Dimensions of the Learning Organization Questionnaire, CA: Jossey-Bass, San Francisco.

Yang, B., Watkins, K., & Marsick, V. (2004) The construct of the learning organization: Dimensions, measurement, and validation. *Human Resource Development Quarterly*. Vol. 15, no. 1, 31-55.

Zhang, D., Zhang, Z., & Yang, B. (2003). Can Chinese state-owned enterprises become learning organizations ? In P.K. Kuchinke (Ed.), *Academy for Human Resource Development Conference Proceedings*, Raleigh-Durham, NC: AHRD Press.

High performance work systems and organizational performance in multinational subsidiaries operating in the Balkan region

Andreea Teodora Balan¹, Alexandros Psychogios², and Athina Vasilaki²,

¹ South-East European Research Center (SEERC), University of Sheffield, anbalan@seerc.org

² Department of Business Administration, City College the international Faculty of the University of Sheffield, a.psychogios@city.academic.gr, a.vasilaki@city.academic.gr

Abstract.

The present study examines the relationship and interaction organizational performance between systems within multinational organization subsidiaries in post communist Balkan countries. The study focuses in finding if (H1) High-Performance Working Practices (HPWS) can be predicted from financial performance, organizational culture, and job satisfaction. The second model (H2) examines if financial performance can be predicted from HPWS, job satisfaction, organizational culture, and organizational citizenship behavior (OCB). Method: Ouantitative data was collected from 181 participants working in MNC subsidiaries operating in the Balkan region (Romania, Serbia, Bulgaria, Albania, FYROM, and Kosovo). The companies were found in the National Management Associations of each participating country. The scale used included items of HPWS, job satisfaction, OCB, Financial performance, organizational culture, as well as demographics. The study used correlation matrices and multiple regression analysis to examine the models. **Results:** Findings showed that HPWS can be predicted from financial performance and organizational culture. Job satisfaction showed no statistical significance in the model. The second model showed that financial performance can be predicted from HPWS and job satisfaction. OCB and organizational culture showed no statistical significance.

Keywords: *HPWS, MNC, subsidiary, organizational performance, job satisfaction, OCB, Balkan region.*

1 Introduction

The use of HR practices to enhance organizational performance has been extensively researched in the fields of organizational psychology, HRM, and industrial relations (Apostori et al., 2008; Takeuchi et al., 2007). Research in the area has mainly focused on individual HR practices and their impact on organizational performance (Fey et al., 2009). In recent years a high amount of attention has been given to the impact of HRM implementation on the employee's psychological wellbeing and performance (Shape & Redman, 2011). In particular, the notion of high performance work systems (HPWS) enjoyed extensive attention over the last years. A plethora of studies has focused on the relationship between HPWS and organizational or employee performance (eg: Butts et al., 2009; Kintara et al., 2006; Liao et al., 2009; Tsai, 2006). However, few studies focused on a model based relationship between HPWS and different HR practices (Butts et al., 2009; Liao et al., 2009), with a large number focusing on individual practices. Furthermore, the high majority of studies has been conducted in western countries (Gittel et al., 2010) or Asian rapidly emerging markets (Liao et al., 2009; Takeuchi et al., 2007). There is virtually a lack of research conducted in East Europe in general and the Balkan region in particular. In order to address the gap in literature, the present study focuses on the HPWS implementation and its relationship with organizational performance in MNE subsidiaries operating in the Balkan region. Specifically the study assesses the link between HPWS, management style, job satisfaction and financial organizational performance in MNE subsidiaries operating in 6 emergent markets in the Balkan region (Romania, Bulgaria, Serbia, Kosovo, Albania, and Skopje).

2. Literature review

2.1 HPWS

Recently, there is an increase interest in HPWS efficiency and its organizational performance link (Becker and Huselid, 1998; Beltran-Martin et al., 2008; Liao et al., 2009). High performance working systems is a set of non-financial performance tools and practices aiming at increasing individual and organizational performance by motivating the individual to internalize the organizational goals (Beltran-Martin et al., 2008). HPWS aims at achieving competitive advantage rather than simply cost cutting by investing in the individual. HWPS refers to the interaction between a set of different HR practices rather than focusing on isolated HR practices. This makes HPWS difficult to define and put into practice. There is a lack of common understanding in the literature to the nature of HR practices involves in HPWS (Fey et al., 2008; Liao et al., 2009). Furthermore, there is a lack of understanding on how to define and name the concept. For example, a number of studies use the notion of high involvement working systems in an attempt to focus on the individual and his/her role

in the organization. High involvement theory advocates argue that employees should be assigned a high level of organizational responsibility and be actively involved in the organization decision making process (Liao et al., 2009). A different school of thought makes use of the term High performance working practices or systems (HPWS) emphasizing on the organizational performance as the final goal in HR practices adoption (Gitell et al., 2009; Beltran-Marti et al., 2008).

There is a plethora of research focusing on the benefits of organizational citizenship behavior (OCB) (Organ, 1997). However, most studies focus on OCB at the employee individual level rather than at the collective organizational level (Cong, Chang, & Cheung, 2010). The notion of collective OCB refers to the work behavior at the group level. A research study conducted on a sample of 239 executive and middle managers in China found a positive association between HPWS and collective OCB (Cong, Chang, & Cheung, 2010). Similarly to the differences in definitions, there is a lack of consensus on what HR practices should be incorporated in the model. For example, the high involvement school maintains that the influence is on the individual and how the model influences him/her (Boxall and Macky, 2009). On the other hand the HPWS advocates focus on the overall organizational performance while still involving the individual. Both schools of thought place a great emphasis on employee empowerement and motivation as the main practices involved in the model (Butts et al., 2009). However, it is important to note that employee empowerment and motivation are abstract concepts, difficult to quantify and open to interpretation differing from organization to organization (Beltran-Martin et al., 2008). The lack of a common definition of HPWS has the potential to lead to confusion when comparing findings from different studies.

The present paper will use the notion of HPWS due to its emphasis on organizational performance while being sensitive to findings of the HIWS research.

A large number of studies linked HPWS with organizational performance (Butt et al., 2009; Fey et al., 2009; Tsai et al., 2010), organizational citizenship behavior (Pare et al., 2007), and organizational support (Butt et al., 2009). In an attempt to analyze the relationship between HPWS and organizational performance, Butts and colleagues (2009) found employees empowerment and organizational support to play a major role in the interaction. By using a large sample of service sector employees from the USA, the study shows that effective HPWS implementation enhances employee empowerment and this in turn leads to higher organizational commitment and overall organizational performance. The study offers an understanding on the mechanism involved in the HPWS and organizational performance link. Research reports a difference in reported HPWS effectiveness between managers and employees (Liao et al., 2009). The study used a sample of middle management and employees working the banking system at 92 world-wide branches of a Japanese Bank. The study found a positive relationship between HPWS and organizational performance as expected from different studies. However, the authors argue that there was a significant difference in reported HPWS efficiency between management and employee as well as intra-departmental differences. Managers had a tendency to over report the relation between HPWS and organizational performance. The study also found within group differences between the employees. Employees are differently exposed to HPWS, thus it is expected to vary in their understanding of HPWS effectiveness. Moreover, the study reported that

the individual's experience of HPWS influenced their job performance (Liao et al., 2009). The findings emphasize the need of managers to focus on the individual differences in HPWS experience. The study should be seen in light of its limitations such as lack of generalization power. Further research reports that HR flexibility acts

such as fack of generalization power. Further research reports that FIX nexibility acts as a mediator factor between the HPWS and organizational performance relationship (Beltrain-Martin et al., 2008). HPWS is defined in terms of selection and recruitment practices, training and development, reward systems, as well as performance appraisal practices. HR flexibility was defines as the individual's ability to effectively gasp new information and use them in accomplishing diverse work tasks. A number of 3427 Spanish organizations commercial managers participating in the study. The results showed a positive relationship between HPWS and organizational performance. However, this relationship losses significance when HR flexibility is introduced in the model. These findings raise the need to further investigate the link between HPWS and organizational performance by adopting a more holistic approach. It is clear that there is a link between HPWS and organizational performance however the question remains on whether there are other factors influencing this relationship and how are they interacting with the model.

2.2 The Balkan Region

Nowadays, there is a plethora of studies focusing on HRM in general and organizational performance in particular (eg: Fey et al., 2008; Gittel et al., 2010; Tsai et al., 2010). However, the high majority of studies are conducted in organizations operating in USA or Asian rapid emerging economies (eg: Takeuki et al., 2007). There is limited amount of research conducted in the Balkan region (Samosi et al., 2011). Furthermore, the existing research focuses on the impact of different HR practices within local organizations, with no studies looking at the impact of HRM in general and HPWS in particular in foreign MNE subsidiaries operating in postcommunist Balkan countries. Regardless of the similar political and historical legacy, the nations in the Balkan region do not have a homogenous culture (Szamosi et al., 2011). Differences include but are not limited to state involvement in the organizational sector, HR policies and state regulations, as well as national culture (Lou ad Tang, 2007). These factors affect the HRM implementation and efficiency in organizations operation in these countries. Until the financial crisis, the region saw a rapid increase in foreign direct investment as well as economical development brought about the increase interest of foreign MNE in the area as well as the privatization of state owned companies.

The region as a whole presents a series of particularities that can be translates into organizational culture in general and HRM in particular. Furthermore, postcommunist states in the Balkan region share a common political legacy than can further impact the national and organizational culture. Former communist countries share similar features such as autocratic and hierarchical management style, limited management-employee communication, state involvement into the organizational culture as well as a high compliance to authority figures (Bedo et al., 2011; Prasnikar et al., 2008). Even though, the iron curtain fell more than twenty years ago and since then the countries have made progress in moving towards a free market economy, the legacy of communist is still present a various degrees having an impact on the business sector (Bedo et al., 2011). A study analyzing the influence of national culture on organizational culture found a strong correlation between the two in three former communist countries (Prasnikar, et al., 2008). Furthermore, the study found Rusia and

Serbia having significantly different national and organizational cultures. Different studies report similar findings in terms of the link between organizational culture and national culture (Sahadev and Demirbag, 2010). One of the main problems of research is the difficulty in finding a case-and-effect rather than a correlation relation between the two.

A large number of MNE subsidiaries operating in the Balkan region, tend to adopt a western HRM system (Brewster et al., 2004). Such an approach can prove problematic due to the cultural and economical differences between the two areas. Western HRM systems might not prove the same success in organizations operating in former communist countries regardless of the organizational culture strength and unity. For example, employees from former communist countries might find it difficult to adopt a direct management-employee communication style due to past legacies such as the highly hierarchical organizational culture (Milikic et al., 2008). Milikic and colleagues (Milikic et al., 2008) further found out that the high majority of organizations operating in Serbia at the time of the interview, had a relatively new HR department with minimal involvement in the organization. The HR departments where mostly focusing on recruitment and selection and monetary reward systems. Similar results were reported (Luca, 2007; Prodan et al., 2009) in organizations operating in Romania. The research conducted in Romanian organizations showed a lack of employee motivational techniques, non-financial reward systems, as well as training and development.

One of the few studies that looked at HR practices implementation as a comparison between former communist countries and capitalist economies, found significant differences between the two groups (Shadev and Demirbag, 2010). The authors found significant differences between the two groups in terms of educational level and employee skills. The former communist countries group had a higher level of employee skills compared to the capitalist countries group. Furthermore, the education level of employees was higher in post-communist countries as opposed to capitalist countries. The study presents a new perspective in the area of HRM. Unlike previous studies, the study uses a set of organizations operating in 8 post-communist countries and 2 capitalist South-East European countries (Turkey and Greece). However, one might wonder if the results would be similar when comparing the postcommunist countries group with a western capitalist group. In a different study, Bedo and colleagues (Bedo et al., 2011) reported a difference in organizational culture between the urban and rural areas. Organizations operating in the metropolitan area benefit from a higher access to information and are more likely to employ a capitalistic oriented HRM system. On the other hand, rural areas are characterized by small and medium organizations who due to the lack of information, tend to have a stricter hierarchical structure.

However, current literature on HRM implementation in the region is scarce and limited in terms of statistical rigor (Szamosi et al., 2011). Furtheremore, there is a limited amount of studies (Bedo et al., 2011; Sahadev and Demirbag, 2010; Szamosi et al., 2011) that are using more than a single country analysis. Due to the differences in the region, it is difficult to generalize findings based on a single country analysis, that being the case of a large amount of studies in the area (Prodan et al., 2009; Svetlik et al., 2010).

2.3 Current study

The purpose of the current study is to explore and analyze the interaction of high performance measurements in a sample of MNC subsidiaries operating in postcommunist Balkan states. The main aim is to predict a model that predicts the effectiveness of HPWS as well as financial performance. The multilevel design will allow to explore the interaction and effects of different organizational performance measurements. Previous research showed a positive relationship between HPWS, organizational financial performance (Tsai et al., 2010), organizational citizenship behavior (Pare et al., 2007), and organizational support (Butt et al., 2009). However, the high majority of research made use of one-dimensional design looking at single interactions. This approached failed to acknowledge the interaction between different practices and measurements. Moreover, a large number of studies come with a series of limitations in terms of statistical methodology adopted (lack of cause-and-effect) and limited study sample. Due to the focus on single countries, result generalization can prove difficult. A limited amount of studies (eg: Galag, 2004; Sahader and Demirbag, 2011; Svetlik et al., 2010) made use of multi-country analysis in the HRM context in general. Furthermore, there is a virtual gap in literature focusing on high performance practices in the Balkan region (Szamosi et al., 2010). Due to the majority of research conducted in western Aglo-Saxon as well as Asian region (eg: Fey et al., 2009; Chen et al., 2005; Takeuchi et al., 2007), there is a lack of understanding on how high performance measurements operate in organizations within the Balkan region. This can have serious practical implications for MNE subsidiaries operating in the region and their effort of introducing HPWS. For this purpose, the current study will use a diverse sample of MNE subsidiaries operating in 6 former communist Balkan countries. The main reason for inclusion was the common cultural legacy brought by the communist area that makes this group of emerging economies an interesting case for analysis.

2.3.1 Hypothesis:

1. The first model will examine the development of HPWS as predicted from levels of financial performance, leadership styles, and job satisfaction. Previous studies focused on correlation analysis failing to assess if one predicts the other. It is hypothesized that high levels of financial performance, job satisfaction, and leadership style will be associated with high levels of HPWS.

The second model will assess the relationship between financial performance and non-financial performance practices. It is hypothesized that financial performance can be predicted from levels of job satisfaction, HPWS, OCB, and leadership style. In other words, high levels of financial performance will be associated with high levels of OCB, job satisfaction, leadership style, and HPWS. In an attempt to better understand organizational performance, recent research found a positive relation between financial and non-financial performance (Katou, 2009). However, there is a lack of understanding in the literature on what non-financial performance measures to look at when assessing the interaction. Few studies found a relationship between financial performance and job satisfaction (Jiang et al., 2011), while other found a

positive relationship between financial performance and leadership style (Hayward, 2005), or HPWS (Beltran-Martin et al., 2008). In the present study it is expected that the different measures of non-financial performance will predict financial performance.

3. **Data and Methodology**

3.1. Design

The study employed primary quantitative data collected between August and November 2011. The sample was recruited from 6 former communist Balkan countries (Albania, Bulgaria, Kosovo, Romania, Serbia, FYROM) using country specific National Management Associations. The language of communication and of the scales used was English language.

3.2. Participants

The study used a sample of 181 participants (females N=82, males N=99) working in MNE subsidiaries operating in former communist Balkan countries. The inclusion criteria were a high command of English language and employee in MNE subsidiaries operating in the 6 post-communist Balkan countries (Albania, Bulgaria, Kosovo, Romania, Serbia, FYROM). An initial number of 250 questionnaires have been received from the organizations. From the initial number of 250 questionnaires received from the organizations, a total number of 69 questionnaires have been excluded from the study. A number of 40 scales have been excluded due to incomplete questionnaires (more than 15%), 7 have been excluded due to working at companies operating outside of the 6 countries analyzed in the study, and 22 participants have been excluded due to low command of English language and inability to understanding the different scale items.

3.3. Materials

All questionnaires were collected using hard-copies. The questionnaires were either collected in enveloped or posted to the researcher.

3.4. Measures

The questionnaire was divided into 4 sections: Management practices, Organizational behavior, Organizational Performance, and Demographics.

Control Measures: Demographic variables included gender, age group, educational level, nationality, country of employment, overall work experience, organizational size, industry, and work experience in the current organization. Predictor variables

The Management practices section includes the high performance working systems scale, and organizational culture scale.

The High performance working systems scale was designed specifically for this study. The scale contains 20 items scored on a 7 point Likert-like scale rated from 1 (not at all) to 7 (great extent). Cronbach alpha was .95.

The Organizational culture scale (Chatterjee, Lubatkin, Schweiger, and Weber, 1992) was used to assess the cultural environment within the organization. The scale includes 8 items scored on a 7 point Likert-like scale rated from 1 (not at all) to 7 (great extent). Cronbach alpha was .85.

The organizational behavior section includes a number of 3 scales: Leadership and Management Style, the Job Satisfaction scale, and the Organizational Citizenship Behavior scale.

The Leadership and Management Style scale (Bass and Avolio, 2004) was used to assess the management involvement style within the organization. The scale had a number of 20 items scored on a 7 point Likert-like scale rated from 1 (not at all) to 7 (great extent). Cronbach alpha was .96.

Job satisfaction was assessed using the General Satisfaction scale of the Minnessota Satisfaction Questionnaire (MSQ) (Arvey and Gross, 1977). The short scale was used to assess the level of employee satisfaction with the current work position. The scale comprises 20 questions assessed on a 7 point Liker-like scale rated from 1(not at all) to 7 (very). Cronbach alpha was .93.

Organizational citizenship behavior (Lee and Allen, 2002) comprises 16 items scored on a 7 point likert-like scale rated from 1 (not at all) to 7 (very). Cronbach alpha was .93.

The Organizational performance section incorporates the Organizational performance scale (Vasilaki, 2011).

Organizational performance (Vasilaki, 2011) was used to assess the perceived organizational financial performance. The scale includes a set of 8 items scored on a 7 point Likert-like scale rated from 1(low satisfaction) to 7 (high satisfaction). Cronbach alpha was .926.

3.5. Procedure

The questionnaires were distributed at 6 different time points in Romania, Serbia, Bulgaria, Albania, and FYROM. The companies were recruited using data from National Management Associations. A first contact was done by phone asking if the organization would be interested to participate in the study.

In July three middle management from MNCs operating in the 6 countries were approached by email, informed of the study's details, and participation was solicited. After their permission of conducting the survey in their company, each company received a batch of questionnaires. The participant was able to fill in the questionnaire at his/her own convenience with the understanding that the deadline would be at the end of August. Due to low response rate the deadline was extended until the end of October with a constant reminder for the participants.

The participants were offered the option of quitting at any point or to receive the completed study after the data analysis.

3.6. Statistical analysis

The questionnaires were checked for missing data. Questionnaires with more than 15% missing data were dismissed from the study.

The data was checked for normal distribution and the missing data was replaced with the items mean.

Factorial analysis was performed as a data reduction tool. Based on factor analysis, 5 items were identified: HPWS scale, Leadership and organizational culture,

Organizational citizenship behavior, Job Satisfaction, and Financial performance. Cronbach alpha was .88.

Correlation matrices and multi linear regression where also performed for the 5 scales.

The analysis was carried out using SPSS 16.0.

4. Empirical Analysis

4.1. Preliminary analysis

The data was checked for missing values and normal distribution tests where performed. Based on this initial analysis, questionnaires with more than 15% missing data have been excluded (N=40). The remaining missing data (less than 15%) was replaced with the mean score of the particular scale. Factor analysis was used as a data reduction technique. Based on the factor analysis, 5 scales have been developed: HPWS (mean= 4.67, sd=1.16), Organizational culture and leadership (mean=4.48, sd=1.04), Organizational citizenship behaviour (mean=5.34, sd=0.85), Job Satisfaction (mean=5.06, sd= 0.99), and Financial performance (mean=4.57, sd=1.20).

The final number of participants included in the study was 181 females (45%) and males (55%). The average educational level was bachelor degree (40.9%) followed by master degree (39%), MBA degree (17%), Phd (1.7%), and other qualifications (0.6%).

The majority of participants had ages between 30 to 35 years old (43%), followed by ages younger than 30 (24%), ages between 36-40 (21%), 41-45 (7%), with the fewest participants being over 45 years old (4%).

Almost half of the individuals had a general work experience of more than 10 years (40%), followed by a work experience of 8-10 years (26%), 4-7 years (21%), 1-3 years (11%), a small percent of participants having less than one year work experience (2%).

The participants had a work experience in the current organization of 4 to 7 years (27%), followed by a work experience of 8 to 10 years (20%), less than a year experience or more than 10 years (16%), and 1 to 3 years experience (15%).

The high majority of the participants were working at the time of the study in the service sector (66.3%), followed by the manufacturing sector (22%), other sectors (6%), and the retailing sector (5%).

In terms of organizational size, more than a half of the participants were working in large organizations with more than 250 employees (54%), followed by small organizations with 20 to 49 employees (21%), organizations with 50 to 99 employees (11%), and organizations with 100 to 249 employees (11%).

The majority of participants had Romanian nationality (42%), Albanian (12.8%), Serbian (12.7%), Skopjan (8%), Bulgarian (8.4%), Kosovar (6%), Greek/Canadian (0.6%), and Austrian (0.6%).

The former communist countries were considered organizations operating in: Romania (43%), Serbia (14%), Kosovo (13%), FYROM (9%), Bulgaria (13%), Albania (6%). (Apendix 1).

4.2. Correlations

Correlation matrices were conducted to identify the correlation between variables. Pair-wise deletion was used.

Table 1: Correlation	matrix (Finan	cial Performance,	Organizationa	l Citizenship
Behaviour (OCB), Hig	gh Performanc	e Working Syster	ns (HPWS), C	Organizational
Culture and Leadership style (Culture/Leadership), Job Satisfaction(JS)):				
1	Ich	Culture / and anothi	- UDWC	OCD

	Job	Culture/Leadership	HPWS	OCB
	Satisfaction			
Financial	.544**	.599**	.615**	.467*
performance	.000	.000	.000	*
Sig. (2-tailed)	181	181	181	.000
Ν				181
OCB	.630**	.615**	.484**	
Sig. (2-tailed)	.000	.000	.000	
N	181	181	181	
HPWS	.600**	.743**		
Sig (2-tailed)	.000	.000		
N	181	181		
Culture/Lead	.757**			
ership	.000			
Sig. (2-tailed) N	181			

** p<0.01 ; * p<0.05

The variables are all positively correlated to one another. Organizational Culture and Leadership style correlated strongly with Job Satisfaction (r=.76), HPWS (r=.74), Organizational Citizenship Behaviour (r=.62), and moderately with financial performance (r=.599). HPWS correlated strongly with Organizational Culture and Leadership style (r=.74), Financial Performance (r=.62), and Job Satisfaction (r=.6). There is a weak correlation between HPWS and Organizational Citizenship Behaviour (r=.48). Financial Performance correlates strongly with HPWS (r=.62), Organizational culture and leadership style (r=.599), moderately with Job Satisfaction (r=.62), and weakly with Organizational Citizenship Behaviour (r=.47).

A collinearity diagnostic was performed to assess the existence of multicollinearity. No evidence of multi-collinearity was found. The tolerance values ranged from .38 to .615. A multi-collinearity would appear when the tolerance values are below 0.1 this indicating a serious problem while the tolerance value below 0.2 should presume a potential problem (Menard, 1995). The VIF scores where between 1.63 and 2.79. In the case of VIF score, a number higher than 10 should be a cause of concern and would require further statistical investigations (Myers, 1990).

4.3. Regression analysis

One of the study hypotheses was that HPWS would be predicted from financial performance, Organizational Culture and Leadership style, and Job Satisfaction. Thus HPWS practices implementation would be associated with high financial performance, high Organizational Culture and Leadership style, as well as high Job Satisfaction. Multiple regression analysis was used to create the model. Using the enter mode of multiple regression, the model showed significance. Adjusted R square=.592, F(2,178)=131.7, p<0.001. Table 2 presents the significant variables of the model. LL A. M. 14' 1

Table 2: Multiple regre Predictors	Beta	р
Financial Performance	.265**	.000

Organizational Culture and .584**

Leadership style

** p<.001; *p<.05

The table shows that the use of HPWS practices can be predicted from the level of Financial Performance, and the level of Organizational Culture and Leadership style. Job satisfaction did not prove to be a significant predictor in the model.

.000

The second hypothesis was that the level of financial performance can be predicted from the level of HPWS implementation, the level of Job Satisfaction, the level of Organizational Culture and Leadership style, as well as the level of Organizational Citizenship Behavior. In other words, it was expected that high level of HPWS implementation together with high levels Job Satisfaction, high levels of Organizational Culture and Leadership style, and high Organizational Citizenship behavior, will positively predict Financial performance.

The model was predicted using enter model multiple regression analysis. The model showed significance at adjusted R squared=.420, F(2,178)=66.09, p<.001. Table 3 presents the significant predictors.

Table 3: Multiple regression Predictor variables	Beta	p
HPWS	.451**	.000
Job Satisfaction	.273**	.000

T 11 2 36 12 1 E. 1 D . . C. 1 1

** p<0.01; *p<0.05

The results show that Financial performance can be predicted from the levels of HPWS implementation and the levels of Job Satisfaction. Organizational Culture and Leadership style as well as Organizational Citizenship Behaviour showed no statistical significance in the model. Thus, high levels of HPWS implementation together with high levels of Job Satisfaction will predict high levels of financial performance.

5. Conclusion

The current study aims at offering an in-depth analysis of the high performance working systems and organizational performance measures. Recent years showed an increase interest in organizational performance with a switch from the traditional financial performance. Non-financial performance in general and HPWS in particular received particular interest. As previously state, research shows a strong link between HPWS and organizational financial performance (eg: Beltrain-Martin et al., 2008). Furthermore, studies reported the interations of factors such as employee empowerement, motivation (Butts et al., 2009), and organizational culture (Boxall and Macky, 2009) in the HPWS organizational financial performance link.

Findings

The first model hypothesized that HPWS will be predicted from organizational financial performance, organizational culture/management style, and job satisfaction. The hypothesis was partially supported. The results showed that financial performance and organizational culture/management style are statistically significant predictors of HPWS. However, job satisfaction failed to show any statistical significance in the model. In the correlation matrix, job satisfaction showed a medium to strong correlation to HPWS (r=.60). However, when the analysis was taken one step further and job satisfaction was introduced into the multiple regression model, the significance was lost. The findings are partially in line with previous studies arguing a relationship between HPWS, organizational culture (Boxal and Macky, 2009), and organizational financial performance (Fey et al., 2009). Contrary to previous studies, job satisfaction showed no interaction in the HPWS model.

The second model analyzed the interaction between organizational financial performance, HPWS, job satisfaction, organizational culture/management style, and organizational citizenship behaviour. We hypothesized that HPWS, job satisfaction, organizational culture/management style, and OCB can predict organizational financial performance. The findings showed that HPWS and job satisfaction are statistically significant predictors of organizational financial performance. Thus, effective HPWS implementation and high levels of job satisfaction can predict high levels of financial performance. Organizational culture/management style and OCB failed to show any statistical significance within the model. In the correlations matrices, organizational financial performance shows a strong correlation with organizational culture/management style (r=.599) and a marginally low correlation with OCB (r=.47). Both items lose statistical significance in the multiple regression model.

When looking at the variable that showed no significance in both regression models, it is important to note that all with the exception of OCB presented medium to strong correlations. Moreover, all variables showed significance when using linear regression while loosing statistical power when introduced in the multiple regression models. A possible explanation of our findings and the discrepancy between them and past literature might lie in the statistical analysis used in the current study. As previously mentioned, a large amount of studies are unidimentional analyzing one interaction at a time and relying on correlations and linear regression analysis. As seen in our study, some variables might show statistical significance in the correlation matrices and linear regression analysis but will lose it when introduced in the multiple regression model. Further research should take into account such differences by employing an interaction model rather than simple correlation or regression analyses.

Practical implications

The current organizational climate in the Balkan region is one of economical difficulty and uncertainty (Milikic et al., 2008). In an attempt to cope in this climate, both local and foreign organizations are trying to employ high performance focused practices that would offer them a competitive advantage. Our study comes in the support of the literature advocating the effectiveness of HPWS in organizational performance. A company centered on the effective implementation of HPWS will show higher employee and organizational performance. However, one should take into account the different variables interacting in the relationship. The present study offers starting platform by creating a model of variables influencing the effective implementation of HPWS as well as the factors influencing organizational financial performance. Furthermore, emerging economies in the Balkan region saw an increase of foreign direct investment as a consequence of the increase in MNE subsidiaries operating in the region (Szamosi et al., 2010). However, due to the lack of literature in the area such organizations have a tendency to either adopt the local organizational culture (Prodan et al., 2001) or implement western HR practices (Szamosi et al., 2010). The present study looks at the HPWS organizational performance model in light of MNE subsidiaries operating in 6 post-communist countries in the Balkan region. The study offers organizations operating in the region a stating point in the efficient adoption of HPWS.

Limitations and Further research

The current study is not lacking a couple of limitations that future studies should be aware of. The organizational financial performance was measured by using a set of self-evaluation items than can be subjective. The study did not use a set of objective financial indications for each organization employed. This might show problematic in light of the diversity of employees participating in the study who might or might not be aware of the organizational financial indicators. The study used both middle managers and employees working in different departments within each company. Thus, it can be feasible to presume that the degree of financial understanding in general and organizational financial performance in particular will vary from department to department and from individual to individual. Furthermore, research shows that management might have a tendency to over report the organizational financial situation (Liao et al., 2009). This might be due to social bias or to protect a secret agenda.

Furthermore, the main assumption of the study is that all participants are currently employed by MNE subsidiaries operating in the region. Even though the selection and sampling procedure are done in such a way to insure that all participants are working in MNE subsidiaries, there might be cases where individuals not working in MNE subsidiaries might have infiltrated in the study. One of the main problem is the scale itself that lacks a question asking participants if they are currently employed in MNE subsidiaries or not. In this case participant failed to report their lack of association with MNE subsidiaries and thus being included in the study even though under normal situations these questionnaires should have been dismissed. However, the possibility is limited due to the strict selection procedure. In the selection and sample procedure only MNE subsidiaries where contacted to participate in the study and the scales where sent specifically to these organizations. Further study should be aware of this and employ a strict selection procedure as well as including an item asking the participant if he/she is working in MNE subsidiaries.

Lastly, the study (communication and questionnaires) has been conducted in English language. This was done in an attempt to avoid limitation accompanying the translation and intra-country comparison problems. The main reason behind it was that English language is a the main communication language in a large amount of multinational organizations (Kaur and Clarke, 2009). However, the study failed to test participants on their linguistic abilities nor did it take into consideration English language exam scores. Thus, there is small possibility that not all organizations presented in the study have a high criterion of English language skills for all their employees.

Regardless of its limitations the present research brings an in-depth insight into the relationship between organizational high performance systems and organizational financial performance in MNE subsidiaries operating in former communist emerging markets in the Balkan region. As previously stated, there is a lack of literature looking at the impact of HPWS implementation in MNE subsidiaries operating in the region. The current study offers a basis for further research in the region. Future studies should attempt a inter-country comparison as well as the employment of causal statistical tests. Furthermore, a comparison on HPWS implementation and organizational performance in foreign MNC subsidiaries and local MNE would offer further understanding in the field.

References:

Apospori, E., Nikandrou, I., Brewster, C., & Papalexandris, N. (2008). HRM and organizational performance in northern and southern Europe. *The International Journal of Human Resource Management*, 7(19), 1187-1207.

Bass, B., & Avolio, B. (2004). MLQ Multifactor Leadership Questionnaire for Research. Redwood City, CA: Mind Garden.

Becker, B.E., & Huselid, M.A. (1998). High Performance Work Systems and firm performance: a synthesis of research and managerial implications. *Research in personnel and Human Resources Management*, 16, 53-101.

Bedo, Z., Demirbag, M., & Wood, G. (2011). Introducing governance and employment relations in Eastern and Central Europe. *Employee Relations*, 4(33), 309-315.

Bedo, Z., Demirbag, M., and Wood, G. (2011). Introducing governance and employement relations in Eastern and Central Europe. *Employee Relations*, 33(4), 309-315.

Beltran-Martin, I., Roca-Puig, V., Escrig-Tena, A., & Bou-Llusar, J.C. (2008). Human Resource flexibility as a mediating variable between high performance work systems and performance. *Journal of Management*, 34, 1009-1046.

Boxall, P., & Macky, K. (2009). Research and theory on high-performance work systems: progressing the high-involvement stream. *Human Resource Management Journal*, 1(19), 3-23.

Boxall, P., and Macky, K. (2009). Research and theory on high-performance work systems: progressing the high-involvement stream. *Human Resource Management Journal*, 19(1), 3-23.

Brewster, C., Mayrhofer, W., Morley, M. 2004. *Human Resource Management in Europe: Evidence of Convergence?*, Elsevier Butterworth-Heinemann.

Chen, S.J., Lawler, J.J., & Bae, J. (2005). Convergence in Human resource systems: a comparison of locally owned and MNC subsidiaries in Taiwan. *Human Resource Management*, 44(3), 237-256

Fey, C.F., Morgulis-Yakushev, S., Park, H.J., & Bjorkman, I. (2009). Opening the black box of the relationship between HRM practices and firm performance. A comparison of MNE subsidiaries in the USA, Finland, and Russia. *Journal of International Business Studies*, 40, 690-712.

Gittell, J.H., Seidner, R., & Wimbush, J. (2009). A relational model of how High-Performance Work Systems work. *Organization Science*,2(21), 1-17.

Gong, Y., Chang, S., and Cheng, S.Y. (2010). High performance work system and collective OCB: a collective social exchange perspective. *Human Resource Management Journal*, 20(2), 119-137.

Jiang, J.Y., Sun, L.Y., & Law, K.S. (2011). Job Satisfaction and organizational structure as moderators of the effects of empowerment on organizational citizenship behavior: a self-consistency and social exchange perspective. *International Journal of Management*, 28(3), 675-693.

Katou, A. (2009). The impact of Human Resource development on organizational performance: test of a causal model, *Institute of Behavioral and Applied Management*.

Kaur, S., & Clarke, C.M. (2009). Analysing the English language needs of human resource staff in multinational companies. *English for specific purposes*, 3(24), 1-10.

Kintara, M.L., Alonso, A.U., & Olaverri, C.G. (2006). High-performance work systems and firms' operational performance: the moderating role of technology. *International Journal of Human Resource Management*, 17(1), 70-85.

Lee, K., & Allen, N. J. (2002). Organizational citizenship behavior and workplace deviance: The role of affect and cognition. *Journal of Applied Psychology*, 87, 131-142.

Liao, H., Toya, K., Lepak, D.P., & Hong, Y. (2009). Do they see eye to eye? Management and employee perspectives of High-Performance Work Systems and influence processes on service quality. *Journal of Applied Psychology*, 94(2), 373-391.

Luo, Y., and Tang, R.L. (2007), International Expansion of Emerging Market Enterprises: A Springboard Perspective, *Journal of International Business Studies*, 38, 481 – 489.

Menard, S. (1995). *Applied logistic regression analysis*. Sage publications. Series: quantitative applications in the social sciences, 106.

Milikic, B.B., Janicijevic, N., & Petkovic, M. (2008). HRM in transition economies: the case of Serbia, *South East European Journal*, 14, 75-88.

Milikic, B.B., Janicijevic, N., & Petkovic, M. (2008). HRM in transition economies: the case of Serbia, *South East European Journal*, 14, 75-88.

Myers, R.H. (1990). *Classical and modern regression application*. 2nd edition. John Wiley. New York.

Organ, D. W. (1997). Organizational citizenship behavior: It's construct cleanup time. *Human Performance*, 10(2), 85-97

Pare, G. & Tremblay, M. (2007). The influence of High-Involvement Human Resources Practices, precedural justice, organizational commitment, and citizenship behaviors on information technology professionals' turnover intentions. *Group* & Organizational Management, 32(3), 326-357.

Prodan, A., Chirlesan, T., & Brailean, I.T. (2001). *Management în administratie* (Management administration), Iasi, Sedcom Libris.

Sahadev, S. and Demirbag, M. (2010). A comparative analysis of employment practices among post-communist and capitalist countries in South Eastern Europe. *Employee Relations*, 32(3), 248-261.

Snape, E., & Redman, T. (2010). HRM practices, organizational citizenship behavior, and performance: a multi-level analysis. *Journal of Managements Studies*, 47(7), 1220-1247.

Stephan, J., and Jindra, B. (2005). Knowledge transfer to MNE subsidiaries based in Central East Europe- Integradting knowledge-based and organizational perspectives. An introduction. *Journal of Economics and Business*, 1,2(8), 135-156.

Sun, L.Y., Aryee, S. and Law, K.S. (2007). 'High-performance human resource practices, citizenship behaviors and organizational performance: a relational perspective'. *Academy of Management Journal*, 50(3), 558–577.

Svetlik, I., Barisic, A.F., Kohont, A., Petkovic, M., Miric, A.A., Slavic, A., Vaupot, Z., & Poor, J. (2010). Human Resource Management in the countries of the former Yugoslavia. *Review of International Comparative Management*, 11(5), 807-833.

Svetlik, I., Barisic, A.F., Kohont, A., Petkovic, M., Miric, A.A., Slavic, A., Vaupot, Z., & Poor, J. (2010). Human Resource Management in the countries of the former Yugoslavia. *Review of International Comparative Management*, 11(5), 807-833.

Szamosi, L., Wilkinson, A., Wood, J., and Psychogios, A. (2010). Developments in HRM in south-eastern Europe. *The International Journal of Human Resource Management*, 21(14), 2521-2528.

Takeuchi, R., Lepak, D.P., Wang, H., & Takeuchi, K. (2007). An empirical examination of the mechanisms mediating between High-Performance Work Systems and the performance of Japanese organizations. *Journal of Applied Psychology*, 4(92), 1069-1083.

Takeuchi, R., Lepak, D.P., Wang, H., & Takeuchi, K. (2007). An empirical examination of the mechanisms mediating between High-Performance Work Systems and the performance of Japanese organizations. *Journal of Applied Psychology*, 4(92), 1069-1083.

Tsai, C.J. (2006). High performance work systems and organizational performance: an empirical study of Taiwan's semiconductor design firms. *International journal of Human Resource Management*, 17(9), 1512-1530.

Tsai, C.J., Edwards, P., & Sengupta, S. (2010). The associations between organizational performance, employee attitudes and human resource management practices. An empirical study of small businesses. *Journal of General Management*, 1(26), 1-20.

Tsai, C.J., Edwards, P., & Sengupta, S. (2010). The associations between organizational performance, employee attitudes and human resource management practices. An empirical study of small businesses. *Journal of General Management*, 1(26), 1-20.

Vasilaki, A., (2011), "The relationship between transformational leadership and post-acquisition performance", *International Studies of Management and Organisation*, 41, 3, 49-68.

Business Relationships between Non- and For Profit Sectors

Maria Magyar¹

¹ Budapest University of Technology and Economics, Department of Management and Corporate Economics, Budapest, Hungary magyar@mvt.bme.hu

Abstract. The purpose of this paper is to provide a better understanding of how network marketing can be used as a marketing strategy and tool to develop business relationships between for- and nonprofit sector. To reach this purpose I investigate the objectives of nonprofit organisations and network marketing nowadays. I also examine the relations between theory, and empirical work in the development of the interaction and industrial network approaches. In my analysis I use the framework developed by researchers of the Industrial Marketing and Purchasing Group. The proposed and potential directions of my work is to present the non-profit sector's break points in the field of sponsorship, cross-border and inter-organisational relationships. This paper presents a short description of different marketing schools' contribution to understand business relationships in the mentioned sectors.

Keywords: nonprofit sector, business relationship, sponsorship

1 Introduction

Business relationship is an important phenomena of business world. Managers have to recognize the significance of business relationships. Business relationships are both opportunities and problems. Ultimately the success of the companies relies on their human resources and business and marketing strategies which can help them to show up in global market. Network marketing has become an increasingly used practice in order to enhance small and medium sized companies' performance. But what about nonprofit sector? Are there any possibilities to help the for- and non profit sectors' cooperation using the network marketing practice?

Marketing is the science of market behaviour. Or to be more precise the scope of marketing is the behaviour of different market actors, as individuals and organizations, suppliers and buyers, middlemen, and market regulating authorities.

Transactions between or among partners are the basic research unites of marketing. Recurrence of transactions creates market dynamics. Market relationships are constructed by repeating transactions (Mandják, 2005). Business market means a series of interactive relationships among mutually dependent and embedded economic and social actors exchanging resources and linking activities. Direct and indirect relationships between the actors create and develop business networks. The main goal of inter organizational exchange is to assure the necessary resources for the organizations proper value creation process.

On the other hand companies have long been involved in community affairs in partnership with nonprofit organizations. But corporate relationships with nonprofit organizations and corporate involvement in public-problem solving have broadened and deepened in both content and form in recent years. This development reflects a new strategic approach to philanthropy and community involvement on the part of a number of corporations. This approach seeks to position corporate philanthropy and community involvement at the intersection of business interest and societal need, tying corporate community activities much more securely to business objectives and responding creatively to the very real limits on corporate philantropic resources in the present competitive environment. Not only are some companies applying this approach to their philantropic activities, they are extendingit to their business activities as well, adopting a concept of corporate citizenship that applies a given set of core values to the way they earn their income as well as how they give part of it away. In the process, these companies are more successfully retaining valued employees, attracting dedicated customers, and fostering public support.

The purpose of my PhD research is to provide a better understanding of how network marketing can be used as a marketing strategy and tool to develop business relationships between for- and nonprofit sector. To reach this purpose I investigate the relations between theory and empirical work in the development of the interaction and industrial network approaches. I also present some details in form of studies about corporate donations in Hungary with aim to describe and prove the companies benefits from its relationships with forprofit organizations.

2 Relationships Between Non- and Forprofit Organizations

The theoretical area of my research involves strategic alliances, cooperative relationships and corporate citizenships. Hence in this section, I breifly review alliance literature and discuss different forms of organizational relationships and the meaning of corporate citizenship.

According to Arndt (1979) there is a tendency of firms engaged in business-tobusiness marketing to develop long-lasting relationships with their key customers and key suppliers rather than focusing on discrete exchanges and has termed this phenomenon ,,domesticated markets". Dwyer, Schurr and Oh in 1987 have pointed out that marketing, defined as building satisfying exchanges, rests on the establishment of relationships. The marketing discipline contains lot of the study of the main aspects and nuances interorganizational and interorganizational relationships has a long and very rich tradition. For instance, findings of Varadarajan and Menon (1988) focused on and have suggested that other manifestations of interorganizational relationships examined in the marketing literature include cause-related marketing (cooperation between a for-profit firm and a not-for-profit institution) and joint sale promotion (cooperation between firms within the confines of the sales promotion component of the marketing mix) according to Varadarajan's (1986) study.

In line with this tradition, marketing scholars have demonstrated a growing attendance in the study of an even broader line of intra- and interorganizational relationships. Confirming the former observation and based on the conceptualization of Morgan and Hunt (1994), there are ten distinct forms of inter and intraorganizational relationships pertaining to the study of relationship marketing. These ten forms of relationships can be further grouped into four types of partnerships:

SUPPPLIER PARTNERSHIPS with (1) goods suppliers and (2) service suppliers;

BUYER PARTNERSHIPS with (6) ultimate customers and (7) intermediate customers;

INTERNAL PARTNERSHIPS with (8) functional departments, (9) employees, and (10) business units; and last but not least

LATERAL PARTNERSHIPS with (3) competitors, (4) nonprofit organizations, and (5) government.

Building on the conceptualization of Morgan and Hunt (1994), strategic alliances represent one form of lateral partnership between competitors. Accoring to this model the supplier and buyer partnerships (1, 2, 6 and 7) indicate vertical relationships between a focal firm and its suppliers and/or buyers. While internal partnerships (8, 9, and 10) deal with horizontal relationships within an organization, lateral partnerships (3, 4- which is the focus of this study and 5) are more concerned with the focal firm's relationships with external parties such as competitors, nonprofit organizations, and government. It is the cooperative relationships between the focal firm and its competitors (3).

On the other hand, Caesar (1986) in his study pointed that the business enhances its public image by being associated with a "worthy cause" and increases its sales in the process. The nonprofit organization receives the cash benefits of the donations, along with the increased public awareness, courtesy of the marketing capabilities of the business.

2.1 Corporate citizenship

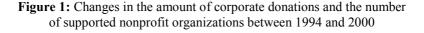
An other interesting and important area of the sector's relationship is corporate engagement and citizenship. One outgrowth of this new approach is a new pattern of corporate engagement with the nonprofit sector, including:

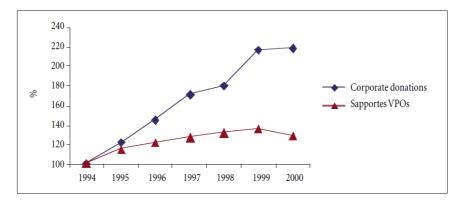
- New forms of giving that go beyond cash and tap the other resources that businesses command, including technical and managerial knowledge and in-kind support;
- New ways to tie philanthropy, and corporate community involvement generally, to the overall business strategies of companies, creating synergies that offer a more secure base for corporate involvement in public problem-solving;

- New types of defining the relationships between companies and nonprofits, with more emphasis on outcomes and deeper corporate involvement in the operations of nonprofit programs that receive corporate support;
- The extension of corporate-nonpfrofit relationships into the commercial sphere through cause-related marketing and sponsorship arrangements (Kotler Keller, 2006). In the former, companies dedicated a portion of their receipts from the sale of particular products to particular nonprofit organizations for use in addressing the problems these organizations are designed to address. In the latter, the corporation provides support for a particular nonprofit activity e.g. a cultural event, a university-based research project in return for public recognition from the nonprofit and, in some cases, joint ownership of the results. Both thus consciously marry the credibility of nonprofit organizations and the public's desire to support them to business marketing goals to achieve beneficial results for society; and
- Broader community partnerships in which companies join with nonprofit organizations, and often government agencies, in long-term, multi-pronged efforts to address complex societal issues, such as the improvement of public schools, and preservation of the environment (The Aspen Institute's Nonprofit Sector Strategy Group, 2001).

3 The Corporate Strategy of Philanthropy

To prove the benefit of companies from its relationships with nonprofit sector I present a brief summary of studies on corporate donations and corporate social responsibility written and published in Hungary by Harsányi-Révész (), Kuti () and Török (). Corporate philanthropy has a long tradition in Hungary, which was not broken in the communist era. For state-run companies it was almost obligatory to develop some form of corporate welfare policy. They had to put part of their profit into a "welfare fund", which then financed corporate welfare services. Several companies had their own nurseries, kindergartens, recreation homes, sports facilities, clubs, libraries and cultural centers (called "culture houses"); most of them regularly supported their old age pensioners and employees in need. This tradition of having a corporate welfare policy did not completely vanish after the advent of privatization. Many firms converted their "welfare funds" into foundations, several corporate welfare institutions were also donated to these foundations before or during the privatization process. These kinds of donations were extremely beneficial for the early development of foundations and accounted for an unusually high share of corporate donations. This share has somewhat decreased since then, but both the amount of donations and the number of supported nonprofit organizations have increased significantly. One quarter of Hungarian NPOs receive some part of their income from corporate donors.





Sources: Nonprofit organizations in Hungary, Annual volumes, Central Statistical Office, Budapest

3.1 Companies' willingness to donate

In 2003, nearly two thirds of Hungarian companies declared themselves to be donors (All the 2003 figures come from a representative interview survey of 1405 corporations.Companies were classified as donors if they said they had provided, without compensation, funding and/or support in-kind to a nonprofit organization in 2003. The companies studied were all "independent legal persons", i.e. companies limited by shares, limited companies, cooperatives, etc. The word "companies" as used in the summary of empirical findings always refers to such entities.) 16 percent said that they do not and definitely would not support nonprofit organizations. One fifth of companies, which did not make any donations, indicated that they might do so in some form in the future. The highest rate of support for nonprofit organizations – nearly four fifths – was among large companies (those with a turnover in excess of 1 billion forints). Considerably less active in their support were small companies (turnover under 20 million forints), of whom only 56 percent made donations and nearly a fifth also ruled out any giving in the future.

3.2 Fields and forms of corporate dinations

Companies provided a total of HUF 31.6 billion to nonprofit organizations in 2003. Of this, nearly 12 billion went to health care. Financial support was also high for nonprofit organizations operating in the fields of social care, education, sports/recreation, and culture. Most benefits in-kind, with some interesting differences in proportion, were also concentrated on these areas.

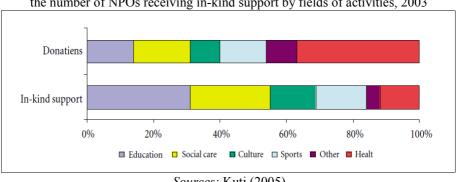


Figure 2: Composition of the amount of donations and the number of NPOs receiving in-kind support by fields of activities, 2003

Over half of the companies which made no donations in 2003 indicated not only their intention to support nonprofit organizations in the future, but also the target group and area of activity to which they would be most willing to direct their donations. Over 40 percent identified helping children (predominantly sick children), as a cause to which they would be likely to make a financial sacrifice. Another quarter identified health and social care (mostly health), as the area they would support in the future.

These indications show a striking parallel, between the priorities of donors and nondonors and also with how the assignable one percent of personal income tax is distributed among different categories. It seems that the owners, executives and managers of companies do not think of donation policy in terms of their company's interests, but as "private individuals". Our interviews with top entrepreneurs also show that they react in the same way to requests for both personal and corporate donations. This implies that their company's first steps towards becoming a donor follow the patterns of traditional charity. Nonprofit organizations in a good position

to win support from them are probably those working in the traditional fields of philanthropy and could be more successful if their requests highlight work among children

The forms of corporate support are varied. As our case studies4 have pointed out, besides the dominant financial support, companies may also:

- . offer services.
- raise the profile of the supported nonprofit organizations, .
- give their own or purchased products to donees, .
- offer awards and fellowships,
- organize voluntary work among their employees,
- facilitate the collection of donations among their employees,
- facilitate the collection of donations among their clients,
- offer cause-related marketing,
- urge their partners to become donors.

These kinds of support are at least as important as and may be even more efficient than financial donations. First, the overwhelming majority of companies offering inkind donations develop closer relations with the supported organizations, become more involved in their work and most frequently also give money. Second, the donor

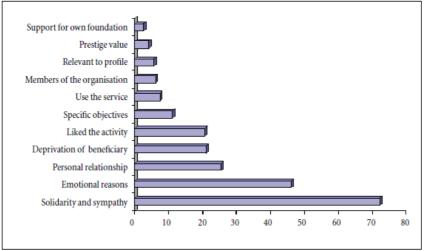
Sources: Kuti (2005)

companies' assistance in solicitation is likely to enlarge the overall amount of donations and to strengthen corporate social responsibility.

3.3 Companies' motivations

According to the results of interviews and empirical survey, social responsibility is the single most important motive for corporate donations in Hungary. The motive mentioned more often than any other – given by nearly four fifths of respondents – was solidarity with, and sympathy for, those in need. Emotional reasons took second place, mentioned by more than half the respondents. A quarter of donors had some personal contact with the nonprofit organizations, which received corporate donations. Clearly, interests were involved here to a certain extent, but there was also a strong emotional background in some cases. A fifth of the interview subjects mentioned the deprivation of the beneficiary organization or the wish to resolve a desperate situation (such as school closure). A similar proportion said that the decision-maker simply liked the activity of the NPO or one of its projects or events.

Figure 3: Number of companies mentioning different reasons for donating as a percentage of all donors, 2003



Sources: Kuti (2005)

4 Future research methods and directions

Eisenhardt (1989) and Dubois and Araujo (2004) in their papers describes case research as one method, has been successfully employed in industrial marketing studies. Their starting point was that a methodology, such as case research, should be seen as a tool that gets appropriated and transformed within the context of a research paradigm. They listed some facts and conclusion why case research method is suitable

to examine the interaction between theory, research methods and empirical work in the development of the interaction and industrial network approaches. In my future work I would certainly take case study into consideration as a possible research method. I will follow Eisenhardt's (1989) process of building theory from case study research. There are many research directions to be followed with the aim to investigate the network marketing's effects in cooperation of the above mentioned sectors. One of the first task is to study the non-profit sector's specialities in the network approach. Further directions of my work is to present the non-profit sector's break points in the field of relationships – sponsorships context (such as Olkkonen-Tuominen in their studies in 2004 and 2005), a case study about successful organisations in the cultural sphere regarding cultural sponsorhip, and cross-border relationships between two or more cultural institutions and their inter-organisational relationships, e.g. relationships between a Hungarian and a Slovak museum and forprofit sector. Further research directions are the network value profile as the marketing orientation of cultural organisations, e.g. museums, galleries, and the different configurations of relationships in cultural-sponsorship relationships in Hungary and Slovakia.

After all I would like to find the answers for the possible research question: Why do firms enter into relationships with the non-profit organizations of cultural sphere, what are the motivating factors, and what are the types of these relationships in practice?

5 Conclusions

New opportunities exist for productive interaction between corporations and nonprofit organizations in solving pressing public problems. Companies are increasingly discovering the importance of community involvement and positive corporate image in order to attract employees, appeal to customers, and secure the license they need to operate in many markets. At the same time, corporations have come to understand that they cannot easily achieve these objectives without the active involvementof nonprofit groups. Nonprofits, for their part, have come to recognize the importance of cooperation with for-profit companies in order to secure needed resources and further validate their activities. Underlying the previously discussed new relationships is a more strategic approach to corporate community involvement. This approach seeks to position corporate philanthropy and community involvement at the intersection of business interest and societal need, tying corporate community activities much more securely to business objectives and responding creatively to the very real limits on corporate philantropic resources in the present competitive environment. To have a meaningful impact, companies thus have to be strategic in their approach, to mobilize the full array of assets that they command, including those of a non-financial character, and to leverage these resources through strategic alliances with other actors. From a particular point of view a large part of corporate donation is fundamentally driven by a sense of solidarity and companies' philanthropic strategy remains basically within the confines of traditional charity. Progress in other fields (such as human rights, culture, community development, etc.), demands more than perfection

of fundraising techniques; there is a need for a change of attitude. A sustained increase in corporate support for civil society issues is only likely if businesses acknowledge the importance of this dimension of socioeconomic development and recognize its implications for their own interests and responsibilities. The best way of engendering and fostering this new attitude would be to replace occasional cash-seeking campaigns with a conscious, professional fund-raising strategy based on collaboration with those involved and to build up sustained contacts between companies and nonprofit organizations based on mutual esteem.

References

Arndt, J. (1979) Toward a concept of domesticated markets. *Journal of Marketing*, 43(4): 69-75.

Caesar, Patricia (1986) Cause-Related Marketing: The New Face of Corporate Philanthropy. *Business and Society Review*, 59 (Fall): 15-19.

Coviello, N.E. and Brodie, R.J. (2001) Contemporary marketing practices of consumer and business-to-business firms: how different are they? *Journal of Business and Industrial Marketing* 16(5): 382–400.

Dubois, A. and Araujo, L. (2004) Research methods in industrial marketing studies. In *Rethinking Marketing. Developing a new understanding of markets, ed.* Dubois, A.and Araujo, 207-227.

Dwyer, F. Robert, Schurr, Paul H. and Oh, Sejo (1987) Developing Buyer-Seller Relationships. *Journal of Marketing* 51:(April): 11-27.

Eisenhardt, Katleen M. (1989) Building theories from case study research. Academy of Management Review 14(4): 532-550.

Harsányi, Eszter and Révész, Éva (2005) Models and practice of corporate donations. *Research Project on Nonprofit Organizations*, Budapest

Kotler, Philip and Keller, K. Lane (2006) Marketing menedzsment, *Akadémiai Kiadó*, Budapest, 922-925.

Kuti, Éva (2005) Corporate Social Responsibility in Hungary. *Research Project on Nonprofit Organizations*, Budapest

Mandják, T. (2005) Néhány gondolat a marketingről és az üzleti kapcsolatokról. *Műhelytanulmányok,* Vállalatgazdaságtan Tanszék: Budapesti Corvinus Egyetem.

Morgan, Robert M. and Shelby, D. Hunt (1994) The Commitment-Trust Theory of Relationship Marketing. *Journal of Marketing*, 58 (July): 20-38.

Nonprofit Sector Strategy Group (2001) The Nonprofit Sector and Business: New vision. New Opportunities. New Challenges. *The Aspen Institute*

Olkkonen, R. and Pekka, T. (2004) Towards an understanding of the fading process of inter-organisational relationships. A longitudinal case study in the context of cultural sponsorhip. *20th Annual IMP Conference*. Copenhagen

Török, Marianna (2005) The practice of corporate donations in Hungary. *Research Project on Nonprofit Organizations*, Budapest

Varadarajan, P. R. (1986) Horizontal cooperative sales promotion: A framework for classification and additional perspective. *Journal of Marketing*, 50(2): 61-73.

Varadarajan, P. Rajan and Menon, Anil (1988) Cause-Related Marketing: A Coalignment of Marketing Strategy and Corporate Philanthropy. *Journal of Marketing*, 52 (July): 58-74.

Keys success factors of New Service Development from the Open innovation perspective. Benefits and implications.

Eriona Shtembari¹ and Gert Pustina¹,

¹ Phd student (Department of Management), University of Tor Vergata, Faculty of Economy, Italy. Email: shtembarieriona@gmail.com

Abstract. The topic is of our interest because open innovation has become an inherent aspect of business nowadays while the research dedicated to innovation in service (hence NSD) is rather limited. This study addresses this gap by exploring the incidence and trends towards open innovation in new service development. The aim is to find out key success factors of NSD in an open innovation context. Our study reveals a list of eight important criteria of success. This consists of three financial criteria: (1) profitability, (2) production cost, (3) return on investment; and five non-financial criteria: (4) strategic fit, (5) marketing criteria, (6) corporate social responsibilities, (7) information quality, and (8) facilitating factors. Furthermore, the benefits and implications of open innovation to NSD are presented in this paper.

Keywords: New Service Development, service innovation, open innovation.

1 Introduction

The service sector makes up more than two-thirds of the economy. The services sector is growing continuously in recent years. More than half of our GDP is accounted from the services sector. This sector dominates with the best jobs, best talent and best incomes. "It is the service elements that make the difference on the marketplace and not because of the product components in the manufacturing's offering" (Grönroos, 1998: 21). Many leading innovative companies are finding that their own businesses are shifting towards services as well (Chesbroug et al. 2011). Innovation is becoming the heart of all businesses. "Companies that do not innovate die" Chesbrough (2003). Moreover, open innovation is changing the way companies are doing business Chesbroug (2006). We have noticed that literature so far lack of research on open innovation and even less research is done on open innovation related to services. But "open innovation can and should apply to services, too" Chesbroug (2011). Therefore, a higher contribution from researchers is needed on the open service innovation.

Moreover, the view that developing new services can benefit from practicing open innovation paradigm is currently even more unknown. With an increasing importance of the service sector, the management of new service development (NSD) is becoming a key competitive concern for many companies (Riedl et al., 2009; Menor, Tatikonda and Sampson, 2002; Johnson, Menor, Roth and Chase, 2000; Fitzsimmons and Fitzsimmons, 2000; Johne and Storey, 1998; Gallouj and Weinstein, 1997). Thus, we argue that a study on key factors fostering successful New Service Development in an open innovation context reveals a gap on the current literature and triggers our research interest. The purpose is to provide practitioners with a comprehensive and practical set of success factors to be taken into account while dealing with NSD projects. This study is conducted with a hope to facilitate the new service development and to add a small part to the current body of knowledge of service management from the view of open innovation.

The research question raising in this paper is:

What are success factors of New Service Development from the perspective of open innovation?

In order to answer the research question, the following research objectives are aimed:

(1) Provide comprehensive understanding of New Service Development by exploring the process, activities, features and implications.

(2) Provide understanding of the new paradigm open innovation and reveal benefits and challenges from the service perspective.

(3) Finally, find out success factors of NSD from the open innovation perspective.

To feel in this research gap and identify the success factors, we apply a systematic literature review on both New Service Development and open innovation. First, in line with the synthesis innovation approach we study NSD concept and created a list of common success factors based on the process, activities, features and implications found from the literature on NSD (Table 1). Second, factors considered as important

for the open innovation process are identified based on general literature on open innovation. Another draft list is established (Table 2). Third, we conduct an empirical study through semi-structured interview approach to verify findings from the literature review. Finally, we conclude with a set of key success factors which are a result of the factors listed in table 1 and 2 combined with findings from empirical study.

With the purpose of providing readers a clear and logical approach to the research topic, the paper is structured in five parts. First, we provide a comprehensive understanding of innovation in services starting from the pioneer contributions of Gershuny and Barras and continuing with the three main current perspectives about innovation in services: assimilation, demarcation and synthesis, and how the latter is gaining in importance nowadays. In the second part we discuss the new paradigm of open innovation coined by Chesbrough (2003) exploring its practices, features and implications. In line with synthesis position, which maintains that services are neither totally different from manufacturing nor share the same characteristics In the third part we concentrate on the situation of New Service Development concept by elaborating on the process, key activities, revealing key features and implications of NSD, by so pointing out important factors influencing NSD. Empirical study is presented in the fourth part. In the fifth part we conclude with relevant findings and provide a set of success factors of NSD from the perspective of open innovation.

2 Literature Review

2.1 Innovation in Services

The service is increasingly contributing to the global economy. According to the US industry statistics, 'the non-good production industry accounts for more than 70% of the total economic activity in United State'. Besides, at the meeting of the OECD Council at Ministerial level, the OECD (2005) reports that service industry gives an important contribution to the growth, productivity and innovation of OECD countries. Other developing countries are also moving towards the service industry instead of the manufacturing one, because of the demand from other countries as well as from their own citizen. Therefore the innovation in service sector is becoming a crucial issue for businesses to compete.

The first analyses about the role of services in innovation appeared only 20 years ago, when authors like Gershuny and Barras indicated the potential impact of the use of new technologies by services. The importance of service sector is further emphasized by researchers (such as Grönroos, 1998; Webster 1994) pointing out that it is the service elements that make the difference on the marketplace and not because of the product components in the manufacturing's offering (Grönroos, 1998: 21). Later on, Coombs and Miles (2000) presented three schools of thought operating in service innovation research, in order to illuminate the differences existing in basic assumptions about service innovation. Gallouj (2002c) states that innovation in service lies across the whole range of learning process such as learning by doing, using, interacting, consulting. Chapman and Soosay (2003) claim that service innovation can be expressed in terms of new service launches. They also say that

innovation in service sector could be both technical and non technical innovation while the latter plays a major role.

Three approaches of service innovation

Like innovation in manufacturing, innovation in services is essentially about change and renewal. One of the most important problem of innovation in services is that has been historically disregarded by researchers because of either their low R&D intensity or patent application. Things are changing and, indeed, we can now identify three approaches for studying service innovation (Coombs and Miles, 2000): the Assimilation, the demarcation and synthesis (e.g Gallouj and Weinstein, 1997; Gadrey and Gallouj, 1998; Djellal and Gallouj, 2000; Sundbo and Gallouj, 2000). According to Gallouj and Windrum (2009), the 'assimilation' approach can be traced back to Pavitt's sectoral taxonomy of innovation (Pavitt 1984). This approach suggests that the study of services can be assimilated within a generic set of theories, originally developed through the study of manufacturing innovation. But, (Tidd et al., 2001) debates that the usual dimensions to describe innovation, which stem from manufacturing, are not fully applicable in a service context due to some differences in service and product innovation. While, the 'demarcation' is the opposite view of assimilation approach. Gallouj and Windrum (2009), mention that demarcation writers argue that completely new, services-specific theories of innovation are required to understand the nature and the dynamics of innovation in services because some forms of innovation are services-specific (e.g Gadrey et al. 1995; Sundbo 1998; den Hertog 2000; Preißl2000). Particular attention is given to the co-production innovation view. This suggests that a distinguishing feature of services is the high degree of interaction that occurs between the user and the service provider. The third study approach is the 'synthesis', which as outlined by Gallouj and Weinstein (1997), Metcalfe (1998), Drejer (2004), and Windrum (2007), seeks to take the recent insights of demarcation writers and to integrate these within insights gained in manufacturing studies within a unifying neo-Schumpeterian framework (organisational, product, market, process and input innovation).

2.2 Open innovation in service

In recent years, open innovation has been changing the way many companies think about developing products. It has become a crucial way of thinking while doing business. Open innovation has been defined as '... the use of purposive inflows and outflows of knowledge to accelerate internal innovation, and expand the markets for external use of innovation, respectively' (Chesbrough et al., 2006).

Riel and Lievens (2003) mention that academic literature (such as Froehle et al., 2000; Storey and Easingwood, 1993; Tax and Stuart, 1997) consider the design of NSD process as internal success factor to provide innovation in a company, while others (Cooper and Kleinschmidt, 2000; Dutta et al., 1999; Storey and Easingwood, 1993, 1996, 1998) see NSD process, in terms of marketing, sales and distribution functions, as external factors related to company innovation.

Chesbroug et al. (2011) has argued for the concept of open innovation that companies should both organize their innovation processes to be more open to external knowledge and ideas and also let more of their ideas and knowledge flow to the outside when not being used internally. Yet initial discussion of open innovation contrasted it to traditional R&D and product development; this had the effect of

placing the focus more on product and technology innovation than on service innovation. Open innovation works somewhat differently in service industry, in part because the role of the customer is different; in a service industry, the customer is often involved in an iterative process with the company that results in a customer experience.

Furthermore Chesbrough et al. (2011) argues that some of the concepts of open innovation apply readily to service innovation. For example, Amazon.com has not only created open service innovation by bringing the "outside in" think of customer reviews on Amazon.com and third parties selling products via Amazon's site but also by taking the "inside out," by, for example, using internal knowledge and infrastructure to create a business selling cloud computing services to other companies.

Features, benefits and implications

Gassman and Enkel (2004) in Gassman, O and Enkel, E, Chesbrough, H (2009), mention that a stable structure entails an accepted system architecture and language, the communication, which is a combination of ideas and technical solutions. Furthermore they argue that supplier involvement can provide buying firms with substantial benefits that range from more "operational" benefits, such as the earlier identification of technical problems, fewer engineering change orders, or the availability of prototypes, to more "strategic" benefits, such as better utilisation of internal resources, access to new or supplementary product and process technologies, reduced technical and financial risks, improved product features, or shorter time to market for new products (Clark, 1989; Birou, Fawcett, 1994; Handfield et al., 1999; Dröge et al., 2000; Ragatz et al., 2002). Some conceptual and empirical evidence hints at the importance of suppliers' innovative capabilities as a major determinant for a collaborative development (Wasti, Liker, 1997; Wynstra et al., 2001; Boutellier, Wagner, 2003; McCutcheon et al., 1997; Handfield et al., 1999).

IP plays a crucial role in open innovation as a result of the in- and outflows of knowledge (Arora, 2002; Chesbrough, 2003, 2006; Lichtenthaler, 2007). Inflows and outflows of knowledge are considered to accelerate internal innovation processes and to better benefit from innovative efforts, respectively (e.g. Chesbrough et al., 2006; Chesbrough and Crowther, 2006). External networking is another important feature which is consistently associated with open innovation (Chesbrough et al., 2006).

De Jong et al. (2009), attempts to explore challenges faced during open innovation process, by reviewing previous studies on this issue: Chesbrough and Crowther (2006) for example identified the not-invented-here (NIH) syndrome and lack of internal commitment as main hampering factors ; when it comes to technology exploitation, leading to the 'only-used-here' (OUH) syndrome (Lichtenthaler and Ernst, 2006). More potential barriers can again be found in the related literature on collaborative innovation. He further ellaborates that Boschma (2005) for example identified various forms of 'proximity' which are essential for effective collaboration. These include cognitive, organizational, cultural and institutional differences between collaboration partners, implying that potential problems may arise due to insufficient knowledge, cultures or modes of organization, or bureaucratic elements. To mention only a few, other potential barriers include lacking resources, free-riding behavior, and problems with contracts (Hoffman and Schlosser, 2001; Mohr and Spekman, 1994).

Summing up, based on the discussions on open innovation features and implications, some important factors have been identified as refer to the Table 1 below:

Table 1. Important factors influencing open inflovation		
Gassman and Enkel (2004)	Company structure, communication,	
	customer involvement, suppliers	
	involvement	
Chesbrough and Crowther (2006)	Company structure and strategy,	
Chesbrough (2003, 2011)	networking, communication, technology	
	exploration/exploitation, IP, customer	
	involvement, Employee involvement	
De Jong et al. (2009)	Technology exploration/exploitation,	
	customer involvement, Employee	
	involvement, venturing	

Table 1: Important factors influencing open innovation

2.3 New Service Development (NSD)

Although, New Service Development (NSD) activities are undertaken often in nowadays business life, it does not exist yet an explicit definition of NSD in literature. Some authors still refer to a broader term which is New Product Development (NPD). NPD, according to Business dictionary is 'the process involved in getting a new product or service to market'. Kelly (2000) defines product development as 'an overall process of strategy, organization, concept generation, product and marketing plan creation and evaluation and commercialization of new product or service'. This concept therefore combines both new product development (in a narrower concept to only manufacturing companies) and new service development. Many articles implicitly mention New Product Development (NPD) or New Service Development (NSD) as innovation in product or service, respectively (OECD, 2000; Menor et al., 2002, Van der Aa and Elfring, 2002). Researchers agree that NSD offers organization's clients not only the changes in product's characteristics but also the changes in process and customer interfaces that their customers have never experienced before. However, Menor et al. (2002) view service innovation is less practical focus than new service development. Thus, the term is still a matter of controversy.

According to Avlonitis (2001) the way of categorizing NSD 'lead to better developed new service and superior performance (p.335). Three most popular classifications categorize NSD in three main groups referring to the degree of innovativeness (Lovelock, 1984), nature of the service sector (Gadrey et al., (1995) and stakeholders's power (Den Hertog, 2000). By combining some types of NSD which have similar characteristics such as the core products that are new to the world or new to the company, all activities or processes that improve or add value to current service are considered under one category.

The importance of studying NSD characteristics was raised by Johne and Storey, since 1998, as they affect organizations' profit most and imply success or failure to the organization. Fitzimmons and Fitzimmons (2000) elaborate on the success factors related to NSD previously discussed by other others, such as positive climate and culture (Cooper and Kleinschmidt, 1995), meeting customer needs (Edgett, 1994),

project synergy is considered as the primarily factor distinguishing success from failure (Cooper and De Brentani, 1991). They stress the configuration of activities as the main success factors related to NSD.

While Gallouj and Djellal (2010). underlines NSD process, with clear pre-planned stages (Cooper and De Brentani, 1991; Cooper and Edgett, 1996) as the most important success factor. Several models of managing NSD process have been built with different number of stages. Researchers and practitioners base their arguments about NSD on product development. Bowers (1987, 1989) in Alam and Perry (2002) developed a eight stage model for NSD process similar to NPD process. Another model of 15 stages was developed by Scheuing and Johnson (1989). Even though this model is more comprehensive than the previous one, it does not address important issues such as cross-functional teams, parallel processing of development stages and cycle time reduction which have been highlighted in NPD models. Therefore, it is said that there is no model of NSD that can match the model of NPD.

Eròcal (2005) identified three factors as common to successful multinational service firms: innovation whether related to process or product, opening up of markets and a motivational work environment, stated in Gallouj and Djellal(2010). Focusing on the "market opening up" factor, the importance of open innovation practices in service is revealed.

A majority of studies discuss the features that differentiate service and manufacturing industries, the key factors that significantly affect the outcomes of NSD. However, the open innovation perspective in NSD was neglected.

Features and implications

The specific features of NSD in regarding to those of NPD were discussed by many authors. Firstly, there was a consensus among literature on the intangibility trait of NSD. Thomas, (1978) stated in De Brentani (1991) that there is a risk on behalf of customers in buying such intangible new service which they could not fully examine the quality prior to purchase. Another issue related to intangible feature of NSD is the difficulty in patterning or property right protection for the new service. Competitors find it very easy to imitate making organization hard to maintain sustainable profit and lessoning innovators' incentive (Easingwood, 1986, Klivans, 1990).

Secondary, the simultaneous production and consumption distinguish service and non service process (Dolfsma, 2004). In service sector, the interaction between suppliers and customers is substantial and therefore the barrier between producing, delivering and purchasing service is blurred (Grönroos, 1990). The relationship between companies and clients is very complex, crucial and lead to long term development of companies as stated by Jackson and Cooper (1988). In additional, De Brentani (1991) recommends that it is better for NSD projects to create more customized than standardized service as it helps organization to adjust quickly to customer's special orders. The above literatures imply that both internal and external communication and interaction is highly crucial in this service industry than that in manufacturing sector.

Moreover, Kelly (2000) realizes that the quality of service and the experience of customers on the service vary every time the service is purchased. This is because of the more human involvement in the transactions in service industry than that of manufacturing. In contradict to De Brentani (1991)'s suggestion above, Lovelock (1983) argues that client might see the quality of service as inconsistence and poor

and perceive the company as unreliable which lead to dramatic issues to the company's development.

Different from products, services cannot be produced in advanced and stocked somewhere before delivering to customers (Gallouj and Weinstein, 1997). Cowell (1988) mentions ownership as the last factor that differentiate NSD from NPD. He argues that customers can only hire or access to service while having the full use of product. According to him, this is the basic difference between these two concepts. However, no practical or managerial implications have been suggested.

In brief, based on the discussions on NSD features and implication, some important factors have been identified as refer to the Table 2.

However, based on the above discussion we do not consider NSD an 'opposed' sector with respect to manufacturing, but rather having shared features and differential characteristics at the same time. This lead us to the synthesis view or, more generally to what has been called, as discussed in Gallouj and Windrum (2009), the 'Rainbow economy' (Coombs and Miles, 2000; Den Hertog and Bilderbeek, 2000; Boden and Miles, 2000; Miles, 2001; Howells, 2000, 2002; Hughes and Wood, 2000). Furthermore, we agree with Windrum (2002, p.19) that 'cathedrals and terrace houses are two very different objects, however, common sets of processes, tools and materials are used to construct these objects. The processes of human innovation are also common, whether expressed in the generation of improved immaterial services or improved manufactured artefacts'. Thus, we should consider the economy as a wide and complex whole of interrelated functions, taking advantage from the open innovation practices.

Source	Important Factors for NSD
Storey and Easingwood (1995) (factors depended on performance measure)	 Sales: effective communication, overall company/product fit, distribution strength, market knowledge, product champion Profitability: staff skills and support, quality of service delivery, compatibility/importance, product/tangible quality, distribution strength, quick response Enhanced opportunity: compatibility/importance, product/tangible quality, market knowledge, product distinctiveness
Shostack (1984b)	 Objectivity, precision, fact- driven, methodologically based
Shneider and Bowen (1984)	 Customer satisfaction
Atuahene Gima (1996b)	 Market orientation Other factors: inter-functional teamwork, product-technology fit
Bowers (1989)	 Involvement of contact personnel and customers
Cooper et al.	 Relationship enhancement

Table 2: Important factors influencing NSD

(1994)	
De Brentani	 Nature of service offerings
(1991)	– Project synergy
	– Market characteristics : Market competitiveness, product
	marker fit,
Edgett and	– Clear defined target market,
Jones (1991)	- Thorough and well organised development process,
	- Effective performance of team members,
	– Sufficient funding,
	– Differentiate service,
	 Top management support,
	 Strong launch campaign
Johne and	- Financial performance (e.g. relative profits to sales;
Storey (1998)	profitability level, and pay-back period)
NSD performance	- Window of opportunity (the degree to which the new product
measurement	opened up new opportunities to the firm in terms of products
	and markets)
	- Market impact (e.g. domestic and foreign market share)

3 Methodology and data

This study is based on data collected through semi-structured interviews. According to Bryan and Bell (2003), in a semi-structured interview, the interviewer has a set of questions on reasonably specific topics to be covered, and the interviewee has a large degree of flexibility in how to reply. This method was appropriate for answering the research question of this study based on the *nature of the question*. The research question about the success factors of open innovation in NSD projects is exploratory and based on the experience and points of view of people involved in the process. Therefore, a qualitative method of research is more suitable for exploring the issues involved in this study.

Triangulation of the data was insured by looking also at secondary data provided by the respondents.

Factors for selecting the interviewees were based on their role in innovation projects and the projects that interviewees deal with must be NSD. The sample included people from different countries such as Albania, Italy, Sweden and Vietnam. Although, they are working in the same service industry, their companies focus on different business area such as financial service, information, airlines, real estate, etc. The reason from choosing a diversity of nationals and sectors is to look at many different perspectives of the topic and enhance the generalization process of the chosen inductive approach. Due to budget constraints, only interviews in Albania were conducted face to face. The purpose for spreading the research focus geographically and including several service sectors is to confirm the findings and to increase the transferability and reliability of the findings.

The questions were formed based on the knowledge gained through literature review and followed closely the purpose of the research. Most questions are open-ended and provide respondents chances to express their views and experiences, which turns out to improve researchers' knowledge on the practical perspective of the topic and contribute to the current body of knowledge.

The interview procedure followed the suggestion that the 'dual approach' should be applied in business and management research (Healey and Rawlinson, 1993 in Bryman and Bell, 2003).First, we made phone calls or sent emails to potential interviewees. Second, a letter of introduction about our research, the purpose and process has been sent to the respondents to follow up the process. Later, a telephone call or email was made a few days after the receipt of the letter to ask for their attendance on the interview process and set appointment. The interviews were carried out on the appropriate time to the interviewees. The length of interview varied from an interviewe to another. However, it is restricted in roundly an hour as time is value to respondents. Many different tools of interview were applied according to interviewee's preferences and time. Some interviews were undertaken in person while some others were carried out via phone or email.

Many difficulties arise prior to the interview process regarding the possibility to approach potential interviewees, especially project managers. Besides, the title 'project manger' or project team are not explicitly used in some non-project based companies. Under such challenges, we interviewed not only project managers of NSD projects but also the team members as well as product managers and other managers who were responsible of launching new services. Since the availability of respondents is crucial for the validity and success of the interview process, personal contacts were used to get in touch with more relevant interviewees.

 Table 3: List of interviewees

N o	Position	Experie nce	Business area	Country based	Interview method	Intervie w time
1.	Country	7 years	D	* **	Email;	50mins
	Product		Banking	Vietnam	Telephone	
2	Manager	5			Emeil.	25
2	IT Manager	5 years			Email; Telephone	35mins
3	Investme	6 years			Telephone	45
5	nt	0 years	Energy/C		relephone	mins
	Council		onsulting	Sweden		mms
	Member		service			
4	Project	4 years			Emails;Tel	50mins
	evaluator	-			ephone	
5	Project	1.5			Email;	61mins
	Manager	years	Air	Sweden	Telephone	
6	Senior	13	navigatio		Email;	30mins
	Project	years	n		Telephone	
	Manager					
7	Product	4.5	Banking	Sweden	Email;	35
0	Manager	years	D 1		Telephone	mins
8	Project	6 years	Real	Albania	Face- to	60
	evaluator		estate		face	mins
9	Consulti	2 years	Financial	Sweden	Face- to	40
	ng		consultin		face	mins
10	manager	2	g	X 7' /	F '1	50
10	Product	3 years	Airlines	Vietnam	Emails;	50
11	Manager Steering	2 100000	Donking	Albania	Telephone Face to	mins 40
11	Steering Committ	2 years	Banking	Albailla	face,	40 mins
	ee				secondary	mms
	Member				data	
12	Expert of	3 years	Research	Albania	Face to	60
	monitori	-	center		face	mins
	ng and					
	controlli					
	ng					
13	PM	2 years	Educatio	Albania	Face to	60
14	PM –	5 years	n service		face	mins
	Sponsor					
	institutio					
15	n Droigat	2	Talasses	Alboria	Emoiler	25
15	Project Manager	3 years	Telecom municati	Albania	Emails; Telephone	35 mins
L	Manager		municati		relephone	mins

			on			
16	Project	2 years	Computi	Italy	Emails;	30
	Manager		ng		Telephone	mins
			service			
17	Project	10	Educatio	Italy	Emails;	45
	manager	years	n service		Telephone	mins

4 Empirical Analysis

All the respondents were kindly asked to mention success factors considered as crucial from their experience with NSD projects. Based on the answers we grouped success factors, according to their nature in two groups: Financial and Non-Financial Factors.

Financial Factors

Factors such as Profitability, Production cost and Return on investment are grouped together based on their financial nature. These factors were investigated and recommended by many authors such as Johne and Storey (1998), Ballantine and Stray (1998), or Müller, (2003). As mentioned by them, the financial aspect of a project is very essential to evaluate projects since it provides numerical idea to project sponsor of how much the project will cost, how much the company will win and for how long the revenue could cover the amount of money invested or how much shareholders gain from the project. These are visible information that could be measured and compare over time. However, these numbers have to be based on realistic data and use suitable projection standard to avoid over or under estimation. Along the project implementation, the data have to be kept updated to see if the project is running on budget or how much changes influence shareholders' value. Respondents commented that *profitability* is not only important to select profitable projects since the early stage of decision making but also during the on-going phase of the project and help assessing the success of the project in monetary terms at the end of project. Half of respondents refer to this criterion as extremely important, five of them said this as fairly important while only three pay less attention on it. The *production cost* in service companies is normally related to technology exploration/exploitation, staffs' salary, promotion cost, etc. Based on the judgement of respondents, this factor is however not much as important as the other two financial criteria as judged by respondents. The explanation for this is due to the difference between service and product development projects. In manufacturing sector, production cost is rather large whereas that in service companies is not significantly a huge investment.

Non Financial Factors

According to Lopes and Flavell (1998), there are other aspects rather than just financial numbers that companies should take into consideration during service innovation such as organizational, managerial, political, social and environmental dimensions. Lack of those criteria could lead to failure of the projects. As mentioned by Alder (2000), non financial criteria related to strategic issues such as 'strategies fit, relationship with stakeholders, competitive advantage, value chain, and cost drive' are very important criteria used to support the decision of selecting or keeping a project from the planning phase until the end of project. In consistence with this, respondents

from all the companies involved in this study mentioned important issues related to strategic fit such as strategy alignment, internal capacity and scope. Although each of them mentioned those criteria in different ways and with various concerns, the results from our interviews support previous study and show that strategic fit is considered among the most important criteria of evaluation. However, the emphasis was given to customer satisfaction aspect. This is consistence with literature such as Den Hertog (2000), Grönroos, (1990) in service that customers' assessment is essential in the development of new service. The results present the critical position of this factor, specifically to new service development projects. Customer satisfaction is considered very important by all respondents from semi-structured interviews. Since customer is the user of the service, their satisfaction is considered one of the best measurements of the quality of service offered through the project. Whereas, quality, which is more sensitive to NSD than to NPD (Tukel and Walter; 2001), defines how successful the NSD project is in terms of market impact. Thus, base on the level of customer satisfaction to the new service, companies could check how much the project fulfilled its objectives and how much impact the service put on the targeted market. To support this, a respondent argued that 'The more satisfy the customers give to the new service during the development of service, the less time needed to redesign and test the new service...It reduces time to market of the service and the duration of the project'. This links to another aspect of marketing criteria- Timing. Although, this factor was not much mentioned explicitly by respondents, it was implicitly referred to as one of the success factors in NSD projects. In support to the recommendation of Farbey et al. (1992), data from semi-structured interviews point out that service companies should take into consideration how much the project match the culture of the company in terms of environmental and CSR issues. Most of respondents agree that company should pay more attention to CSR and consider it as an evaluation criteria as advocate to the trend toward CSR in academic world. However, the result shows that companies are still more concerned on the firms' benefits than to the societies'. The involvement of stakeholders and their attitude toward the new service is considered as facilitating factors (Andersen et al., 2002). This factor was also mentioned by respondents. It could be the attitude of external stakeholders or of internal staffs to the project. However, different from Andersen et al., (2002), NSD practices pay more attention to the involvement of customers to the projects. This could be implied that customers play a crucial role in the success of new service development projects and their participation in the development of new service contribute to the quality and results of projects. Moreover, data collected from semi-structured interviews support the findings of Tukel and Walter (2001) and Steven et al. (1993) that companies should pay attention to the *communication channel* and the *quality of information* gathered as it will affect the quality of innovation process. We noticed that the respondents from companies with many levels of organization are very much concerned with accuracy and adequacy of the information provided from the feedback and feed forward in the communication channel.

This study reveals some benefits from employing open innovation paradigm in the process of NSD. Most of the respondents argued that through their innovation projects they improve organizational learning. The knowledge transfer is what the companies are looking at while dealing with NSD projects even after the implementation.

All the data collected agree on the following benefits:

- (1) To appraise the worthiness of the investment or to prioritise project portfolio.
- (2) To enhance the organizational learning.
- (3) Improve the ability to respond quickly to the market
- (4) Align service quality with customers' expectation overtime

The previous literature also pointed out these objectives of organizations when dealing with NSD projects (Ye and Tiong, 2000; Mohamed and McCowan, 1999; Chapman, 2005).

"Aligning service quality with customers' expectation overtime " was mentioned explicitly in all interviews. We noticed that open innovation has a major contribution at promoting organization learning, which is consistent with the recommendations of IDRC and APM (2006) and emphasise the trend of moving further to ongoing learning rather than just measuring in today's market (Brown and Remenyi, 2002). Knowledge transfer and learning is the basic of open innovation and thus, becoming even more critical to NSD projects. Moreover, our respondents expressed their serious concern to the alignment between service quality and customer's expectation during innovation. As discussed by Kelly (2000), the quality of service and the experience of customers on the service vary every time the service is purchased, it is necessary to verify their expectation all the time to ensure the satisfaction of customer on the service. The fact, however, was not emphasised much by researchers in service innovation. This shows the specific difference between the aim of innovation in NSD and that of NPD.

The cherry on top of the cake is dedicated to customers. The aim is to align service quality with customer's expectation to provide value- added service to them. This implies that in NSD projects, customer's involvement is more critical to the success of project than that in NPD projects.

The data collected from semi-structured interviews revealed several problems while dealing with innovation to develop new services. Some of them are related to the special characteristic of new service while some are associated to the innovation in general. Most of the challenges are in consistence with what have been found by previous studies on both open innovation and new service development. The main difference between literature and practice is related to the tactical concern of academic versus the more dynamic attitude of practitioners.

An airline company operating in Vietnam mentioned *communication channel* as a challenge in their innovation process whereas the Albanian telecommunication company AMC added that they are having difficulties to gather reliable information. This confirmed the argument of McNamara (1994) that sources of information are very important and lack of good communication channel may lead to a wrong results which dramatically affect the decision making procedure. The problem calls for more attention and effort on how to enhance the feedback and feed forward system within companies as well as with external stakeholders in practice. Moreover, there should be a clear verification process to validate the information collected. The willingness of people to participate and share their knowledge allows the innovation results to be more reliable which turn out to ensure the success of NSD projects.

Although companies are trying hard to improve communication mechanism by flatter the company's structure, some multinational or large companies are still considering *company's hierarchy and procedure* as a factor that complicates the innovation and knowledge sharing process in NSD projects. By analyzing our data collected through interviews with several companies in different countries, it came out that in some companies, (such as the banking company in Vietnam), where the decision making is centralized in one regional office, it is not easy to come up with common innovation indicators. Some innovation measurement factors are not agreed among the regional office and local branch resulting in slower the innovation process and influence the quality of the entire NSD project. This problem however was not notified by previous studies.

Even though, all the interviews came into the same position that it is very important to *involve customers* into the innovation process, the challenge associated to the novelty feature of NSD projects is how to get customers involved especially to make them try the service during the development of service. This was also raised by Thomas (1978) and Chesbrough (2011) that customers find it risky to try the new service as they could not fully examine the quality of service prior to purchase. Therefore companies face difficulties to fully engage customers. Another problem related to customer involvement could be inferred that the comments are changed from time to time due to the human trait of customers, so it may take longer time than expected to get correct responses. The implication of this is that customers play a crucial role in securing the quality of new service development projects.

The following table summarizes the problems faced from the companies. It illustrates the consequences in service development and the impact on company's innovation.

No of respo ndent	Source of problem	Consequence	
10	Communication channel	 Wrong measurement results Low stakeholder involvement in innovation process 	
3	Company structure and procedures	 Difficulties on determining indicators Complicated innovation procedure 	
11	Intangibility characteristic	 Difficulties on measurement of prestige, competitiveness, social effects. Subjective measurement 	
17	Customer involvement	Subjective measurement resultsTime consuming	

Table 4:	Challenges	associated	with	innovation	from N	٩SD	projects
----------	------------	------------	------	------------	--------	-----	----------

5 Conclusions

Empirical data provided from interviews show that open innovation concepts apply to New Service Development and close collaboration with customers lies in the heart of NSD process. The justification is that service companies are becoming more and more customer focused and the reason why they develop new services is mainly to fulfil and satisfy customer needs. This finding adds up to the current service management research.

The distinction between the innovation process for NSD and that for NPD is mainly on the intangibility and the human involvement characteristics of service which complicates this process. This provides challenges for practitioners in capturing new ideas and develop new services. It is also hard to get customers involved and obtain their valid comments as their answers are very emotional and inconsistent over the time. Therefore the innovation of NSD projects is usually considered as more subjective.

Furthermore, we found that open innovation has a considerable contribution at promoting organization learning and aligning service quality with customers expectations.

To conclude, from our research through semi structured interviews with several companies in Albania, Italy, Sweden and Vietnam a list of factors is finalized as illustrated in the Table 5 below:

Financial Factors	Non Financial Factors
 Profitability criteria (revenue, number of customer, growth) Production criteria (technology exploration/exploitation, IP, promotion cost, salary) Financial parameter (net present value, return on investment, cost of capital, payback period) 	 Strategic fit (venturing, company structure, objectives, strategy, policies, corporate values, company's capacity) Marketing criteria (customer/supplier involvement, networking, time to market) Environmental and CSR criteria Information quality criteria (accuracy, adequacy, appropriateness of feedbacks and feed forward) Facilitating criteria (stakeholder attitudes and participation)

Table 5: Final list of success factors in NSD projects

We found that the above list of factors (Table 4) combines all the data found from the empirical study together with the contribution of researchers discussed in the literature review. Therefore we can suggest this list of factors as most important to take into consideration during the process for new service development projects. However, different factors have different importance on the service innovation process, depending on the nature and type of service projects.

6 Implications and Limitation of the study

This study calls the attention regarding challenges that interviewed companies are facing during NSD project innovation. The biggest difficulty they have to deal with is customer involvement. As revealed from theory, customer involvement is a very

important criterion for NSD projects unlike NPD projects. Therefore it influences strongly the results of the project.

The main limitation of this study is related to the method applied. One of the drawbacks of the qualitative strategy which is chosen by this study is too subjective, impressionistic making it difficult to generalize findings to other settings (Bryman and Bell, 2003). Besides, since the sample of the study is small makes it difficult to us to generalize some of the findings stated by some respondents. A quantitative method with larger sample size would definitely improve the results of the research.

During the study, we also faced some obstacles when collecting empirical data. Some contacts showed interest to participate in the interviews at the beginning, kept promising that they would be available for the interviews but finally, they did not show up or cancelled the appointment. The reason for this is the industry that we are conducting the study is service sector; most of them are from financial area. The last three months are the very peak time of the year resulting the last minute of refusals by the potential interviewees.

Besides, the time and budget constraints reduce the ability to get more data from the corporate world.

7 Suggestion for further study

Since the study is among the initial efforts to connect service development field with open innovation research, further studies could be approached from either service perspective or open innovation.

From the aspect of NSD projects, this study reveals that there is a concern of project professionals and literature on the management of stakeholders. Since the involvement of customers is of crucial importance during the development of NSD project, it could be interesting to conduct a study on stakeholder management for NSD project.

In addition, open innovation in service underlines the crucial role of close collaboration with customers. Therefore, it could be of particular interest a further research regarding solutions on how to get customers more involved in service innovation.

Otherwise, the same research topic would be further investigated by applying restrictions to geographical location or to business area.

References

Adler, R. W. (2000). Strategic Investment Decision Appraisal Techniques: The Old and the New. *Business Horizons*, 43 (6), 15-22.

Alam, I. & Perry C. (2002). A customer-oriented new service development process. *Journal of Service Marketing*, 16 (6), 515-534.

APM. (2006). APM Body of Knowledge. Buckinghamshire: Ingmar Folkmans.

Atuahen-Gima, K. (1996a). Differential potency of factors affecting innovation performance in manufacturing and services firms in Australia. *Journal of Product innovation management*, 13 (1), 35-52.

Atuahene-Gima, K. (1996b). Market orientation and innovation. *Journal of Business*, 35(2), 93-103.

Avlonitis, G. J., Papastathopoulou, P. G., & Gounaris, S. P. (2001). An empiricallybased typology of product innovativeness for new financial services: Success and failure scenarios. *The Journal of Product Innovation Management*, 18 (5), 324-342.

Ballantine, J., & Stray, S. (1998). Financial appraisal and the IS/IT investment decision making process. *Journal of Information Technology*, 13 (1), 3-14.

Berry, L. (1980). Services Marketing Is Different. *The Journal of Business*, 30 (3), 24-29.

Bowers, M. (1989). Developing new services: improving the process makes it better. *Journal of services Marketing*, 3 (1), 15-20.

Bryman, A., & Bell, E., (2003). *Business Research Methods*. New York: Oxford University Press Inc.

Chapman, R.L., & Soosay, C., (2003). Innovation in logistic services and the new business model A conceptual framework. *International Journal of Physical Distribution & Logistics Management*, 33 (7), 630-650.

Chesbrough, H.W. (2003) *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Cambridge, MA: Harvard Business School Publishing.

Chesbrough, H.W. (2006) *Open Business Models: How to Thrive in the New Innovation* Landscape. Cambridge, MA: Harvard Business School Publishing.

Chesbrough (2006). *Open innovation: the new imperative for creating and profiting from technology*. Harvard Business Press.

Chesbrough, H.W. (2007) *Why companies should have open business models*. MIT Sloan Management Review, 48, 2, 22–28.

Chesbrough (2011). Open Services Innovation: Rethinking Your Business to Grow and Compete in a New Era. John Wiley and Sons.

Chesbrough (2011). Bringing open innovation to services. MIT Sloan Management Review, 52 (2), 85-90.

Citigroup website: http://www.citigroup.com

Coombs, R. and Miles, I. (2000) Innovation, measurement and services: the new problematique. In J.S.

Metcalfe and I. Miles (eds), Innovation Systems in the Service Economy: Measurement and Case Study Analysis, London: Kluwer Academic Publishing, pp.85–104.

Cooper, R., Easingwood, C., Edgett, S., Kleinschmidt, E., & Storey, E.C. (1994). What distinguishes the top performing new products in financial services. *Journal of product innovation management*, 11 (4), 281-299.

Cowell, D. W. (1988). New Service Development. *Journal of marketing management*, 3 (3), 296-312.

Creswell J.W (2003). *Research Design: Qualitative, Quantitative, and Mixed Method Approaches*, second edition. Sage Publications. Inc.

De Brentani, U. D. (1995). New Industrial Service Development: Scenarios for Success and Failure. *Journal of Business Research*, 32 (2), 93-103.

De Brentani, U. D. (1991). Success factors in developing new business services. *European Journal of Marketing* 25 (2), 33-57.

De Jong et al. (2009), Open innovation in SMEs: Trends, motives and management challenges. *Technovation 29, 423-437*.

Dolfsma, W. (2004). The Process of New Service Development - Issues of formalitization and appropriability. *Erasmus research institute of manageement*, 1-20. Drejer, I. (2004). Identifying innovation in surveys of services: a Schumpeterian perspective. *Research Policy*- Elsevier

Easingwood, C. J. (1986). New Product Development for Service Companies. *Journal of Product Innovation Management*, 3 (4), 264-275.

Edgett, S., & Jones, S. (1991). New product development in the financial service industry: a case study. *Journal of marketing management*, 7 (3), 271-284.

EON website: http://www.eon.com

Fitzimmons, J. and Fitzimmons, M (2000). New service development: creating memorable experiences. Sage Publications.

Gassman, O and Enkel, E, Chesbrough, H (2009). Open R&D and open innovation: exploring the phenomenon. *R&D Management. Volume: 39, Issue: 4, Publisher: Wiley Online Library, Pages: 311-316*

Gadrey, J., Gallouj, F., & Weinstein, O. (1995). New modes of innovation How services benefit industry. *International Journal of Service Industry Management*, 6 (3), 4-16.

Gallouj, F., & Weinstein, O. (1997). Innovation in services. *Research policy*, 26 (4-5), 537-556.

Gallouj, F. (2002c) Innovation in the Service Economy. The New Wealth of the Nations, Cheltenham:Edward Elgar.

Djellal and Gallouj, (2008). *Measuring and improving productivity in services: issues, strategies and challenges*. Edward Elgar Publishing Ltd. UK.

Gallouj and Windrum (2009). Services and services innovation. *Journal of Evolutionary Economics, Springer*.

Gallouj and Djellal(2010). The handbook of innovation and services: a multidisciplinary perspective. Edward Elgar Publishing Ltd. UK.

Griffin, A., (1997). PDMA research on new product development practices: updating trends and benchmarking best practices. *Journal of Product Innovation Management*, 14 (6), 429–458.

Grönroos, C. (1990). Service Management and Marketing: Managing the Moments of Truth in Service Competition. Lexington: Lexington Books, Lexington MA.

Grönroos R. (2000). Service Management and Marketing. A customer relationship management approach. Second edition. John Wiley and Sons, ltd, England

Grönroos R. (1998). *Service marketing theory: back to basics*. Working paper. CERS, Finland.

Hertog, P. (2000). Knowledge-Intensive Business Services as Co-Producers of Innovation. *International Journal of Innovation Management*, December, 4(4), 491.

International Development Research Centre (IDRC), Canada. http://www.idrc.ca

Jackson. R. W,. & Cooper. P. D., (1988). Unique Aspects of Marketing Industrial Services. *Industrial Marketing Management* 17 (2), 111–118.

Johne, A., & Storey, E.C., (1998). New service development : a review of the literature and annotated bibliography. *European Journal of Marketing*, 32 (3), 184-251.

Johnson, S.P., Menor, L.J., Roth, A.V., & Chase, R.B., (2000). A critical evaluation of the new service development process: integrating service innovation and service design. Sage Publications, Thousand Oaks, CA, 1–32.

Jong, J. d., Bruins, A., Dolfsma, W., & Meijaard, J., (2003). Innovation in service firms explored: what, how and why? Literature review. *Business and Policy Research*, Available at: http://www.ondernemerschap.nl/pdf-ez/B200205.pdf (Accessed online 01November 2008)

Jong J. & Vermeulen P., (2003). Organising successful new service development: a literatue review. *Management Decision Journal*, 41 (9), 844-855.

Kelly, D., (2000). New service development: initiation strategies. *International Journal of Service Industry Management*, 11 (1), 45-62.

Klivans J. M., (1990). Launching a financial service: A case study in persistence. *The Journal of business strategy*, 11 (5) 8-11.

Langeard, E. & Eiglier, P., (1983). 'Strategic management of service development', In Berry, L.L. *et al.* (Eds), *Emerging Perspectives on Services*, AMA, Chicago, IL, 68-72.

Levitt, T., (1976). The industrialization of service. *Harvard Business Review*. 54(5): 63-74.

Lopes, M. D., & Flavell, R. (1998). Project appraisal a framework to assess non-financial aspects of projects during the project life cycle. *International Journal of Project Management*, 16 (4), 223-233.

Lovelock, C. (1984). *Developing and implementing new services*. Chicago: American Marketing Association.

Lovelock, C. H. (1983). Classifying Services to Gain Strategic Marketing Insights. *Journal of Marketing*, 47(3), 9-20.

Luftfartsverket Group website: http://www.lfv.se

Maister, D.H., & Lovelock. C. H, (Summer 1982). Managing Facilitator Services. *MIT Sloan Management Review*. 23 (Summer 1982),19–31.

McNamara J.F (1994). Surveys and experiments in education research. Lancaster. PA, Technomic.

Menor, L. J., Tatikonda, M. V., & Sampson, S. E. (2002). New service development: areas for exploitation and exploration. *Journal of Operations Management*, 20 (2), 135-157.

Menor, L.J., & Roth, A.V., (2007). New service development competence in retail banking: Construct development and measurement validation. *Journal of Operations Management*, 25 (4), 825–846.

Meyer, M.H & De Tore, A. (2001). Perspective: creating a platform-based approach for developing new services. *Journal of Product Innovation Management*, 18(3), 188-204.

Müller, R. (2003). Commercial Aspects of Project Management, An Introduction. *Umeå School of Business and Administration, Umeå University*, 1-14.

OECD. (2000). Promoting innovation and growth in services. Organisation for Economic Cooperation and Development: Paris.

Riel A. & Lievens A. (2003). New service development in high tech sectors: a decision making perspective. In Van Riel, Allard C.R (2002), *Effective Decision Making in the High Tech Service Innovation Process*, Doctoral Dissertation, Maastricht University, Maastricht, Datawyse/Maastricht University Press.

Saunders, M., Lewis, P., & Thornhill, A., 2007. *Research Methods for Business Students*. England: Pearson Education Limited.

Scheuing, E.E & Johnson, M.E. (1989). New product development and management in financial institutions. *International Journal of Bank Marketing*, 7 (2), 17-21.

Schneider, B., & Bowen, D.E., (1984). New service design, development and implementation and the employee. In George, W.R., & Marshall, C.E. (Ed.), *Developing New Services*, American Marketing Association, Chicago, IL, 82-101.

Shostack, G. L. (1984a). Designing Services that Deliver. *Harvard Business Review* 62 (January-Febuary), 133-139.

Shostack, G.L. (1984b). 'Service design in the operating environment', in George, W.R. and Marshall, C.E. (Eds), Developing New Services, American Marketing Association, Chicago, IL, 27-43.

Storey E.C. (July 1993). The Impact of the New Product Development Project on the Success of Fmancial Services. *The Service Industries Journal*, 13 (3), 40-54.

Storey, E.C., & Easingwood, C. (1995). Determinants of new product performance: a study in the financial services sector. *International Journal of service industry management* 7 (1), 32-55.

Storey, E.C. & Easingwood, C. (1994). New service success and the augmented service offering. Proceedings of Product Development & Management Association 18th International Conference, Boston, 5-10 Nov. 178-89.

Storey C. & Kelly D. (2001) Measuring the Performance of New Service Development Activities. *The Service Industries Journal*, 21 (2),71–90.

Strauss A. & Corbin J. (1990). Basics of Qualitative Research. Techniques and Procedures for developing Grounded Theory, second edition. New Burry Park, CA: Sage Publications Inc.

Suwardy T., Ratnatunga J., Sohal A.S., & Speight G. (2003). IT projects: evaluation outcomes and impediments. *Benchmarking: An International Journal, 10 (4) General review*.

Tidd, J, et al. (2001). *Managing Innovation - Integrating Technological, Market and Organizational Change* (2nd edn), Chichester: John Wiley & Sons.

Thomas, D.R.E. (1978). Strategy in different in service businesses. *Harvard Business Review*, July-August, 158-67.

Tukel O. I & Walter O. R (2001). An empirical investigation of project evaluation criteria. *International Journal of Operations & Production Management*, 21 (3). 400–416.

Van der Aa, W. a. (2002). Realizing innovation in services. *Scandinavian Journal of Management*, 18 (2) 155-171.

Valeri & Rozenfeld (2004). Improving the flexibility of New Product Development (NPD) through a new quality gate approach. *Society for Design and Process Science*, 8 (3), 17-36.

Windrum, P. (2007). *Innovation in services*. In H. Hanusch and A. Pyka, eds. The Edward Elgar companion to neo-Schumpeterian economics. Cheltenham: Edward Elgar

Wholey J, Hatry H, Newcomer K. 1994. *Handbook of Practical Program Evaluation*. Jossey-Bass: San Francisco, CA.

Determining the Impact of Interorganizational Relationships (IOR) on Innovation: Theoretical Framework Building

Danimir Ljepava¹

¹ South East European Research Centre (SEERC), the University of Sheffield

Email: daljepava@seerc.org

Abstract. The main issue under investigation was to understand innovation outcome as it is influenced by interorganizational (IOR) between companies and relationships external organizations and entities in technology service companies. Methodology of desk research was used in development of a theoretical framework connecting innovation and interorganizational theories, as a clearly identified knowledge gap in the academic literature. Research observes innovation outcome through multiple dimensions of interorganizational relationships (type of relationships, partner type, position within a network, strength of relationships, social aspects of relationships, interorganizational knowledge transfer et sim.). Findings indicate that one of the most important aspects of IORs positively influencing innovation outcome is interorganizational knowledge transfer. Technology transfer from developed to developing countries – such is SEE – seems to positively influence creation of innovations novel to the recipient country. Originality of this research is in providing new views in connecting innovation and interorganizational theories. Importance of understanding the influence of these relationships innovation performance is significant to corporate to performance, business strategy and economic growth in both developed and developing countries.

Keywords: *innovation, interorganizational, relationships, knowledge, transfer*

1. Introduction

Purpose of present research is to further the knowledge on innovation outcome of technology service sector companies as it is influenced by interorganizational relationships (IOR) amongst companies and external organizations and entities. These relationships are also observed as an interaction between organizations in developed countries and developing countries in SEE. This study attempts to provide a theoretical framework connecting innovation and interorganizational theories, as it is identified as a knowledge gap in the academic literature.

1.1 Significance of the topic

The survival in today's challenging global environment is only possible if companies are capable of evolving into dynamic and flexible organizations that can rapidly adapt to new environmental circumstances, and if they can serve continuously changing needs of their global customers (Fallah and Lechler, 2008). It seems that companies no longer have a choice whether they should innovate or not. The consequence of not innovating is that a global competitor, regardless of its size or origin, will have a chance of competing against a non-innovative company and taking away its market share if they come up with a better product or service. As such, non-innovative companies simply cannot sustain forces of the global competition (Goyal and Pitt, 2007). Innovation and especially radical innovation is a very powerful force that can create new markets and new growth (Gilbert, 2003). Innovation is a core microeconomic driver of macroeconomic growth (Clayton *et al.*, 2001), in turn rewarding companies who innovate have a larger market shares and are more profitable comparies who innovators (Fallah and Lechler, 2008; Leiponen, 2000).

External organizational relationships are very important for a company, as no single organization can survive for a prolonged period of time on its own in today's global environment (Fallah and Lechler, 2008; Mrinalini and Nath, 2008; Goval and Pitt, 2007). This network of organization and interorganizational interactions is the heart of how industries are organized (Davis and Eisenhardt, 2007). Competition today is no longer a competition between individual companies, but it is a competition between ecosystems (Moore, 1993). As such, it is imperative for an organization to position itself within its industry ecosystem in the most effective way, not only from the survival, but also from the innovation point of view (Phelps, 2010; Bergman and Maier, 2009). Innovation performance of an individual organization seems to be in a close relationship with interorganizational interactions as collaborative relationships amongst organizations are found to positively influence innovation performance (Palmatier, R. et al., 2007; Davis and Eisenhardt, 2007; Faems et al., 2005). This is the reason why interorganizational relationships and their influence to innovation outcome were chosen as a topic of this study, rather than observing innovation at the level of an individual company.

The most dynamic interorganizational relationships seem to be found amongst technology service companies who have the highest industrial innovation activity (Davis and Eisenhardt, 2007). Further, the service sector contributes to a major portion of GDP in developed countries (Burgess, 2011; Tether, 2005). Investments in the service sector can also produce an influential growth in developing countries, such is SEE (Sen, 2011). Therefore, as innovation efforts within the technology companies seem to have the most dynamic interorganizational relationships and also as the service sector seems to influence a larger portion of the economic growth in both developed and developing countries, this is the reason as why technology service companies where chosen as a further industrial focus of this study.

2. Literature Review

In forming interorganizational relationships, parties entering such an arrangement have certain motivations for existence of the relationship. From the perspective of organizational economic theory, interorganizational relationships are created when a specific activity can be conducted in a more efficient way when performed in a partner relationship then when conducted separately by an individual company (Parmigiani and Rivera-Santos, 2011). This is in line with theories of networked innovation, describing a group of companies working together on innovating new products and services, however sharing each other's resources, risks and rewards - as several companies can provide much larger resources than a single one, innovation capacities of the group are greatly larger than of a single company innovating by itself (Andersen and Drejer, 2008). Increasing innovation capacity through having access to a larger poll of resource is also supported by an RBV¹ being an internal driver of innovation arguing that main internal drivers of innovation is focus on increasing company's financial performance, product quality and new product success (Paladino, 2007).

Viewing motivators for existence of interorganizational relationships, from the perspective of organization theory companies establish partnerships in order to increase effectiveness of different tasks and reinforce interorganizational and interpersonal relationships (Parmigiani and Rivera-Santos, 2011). Such motivator for increasing effectiveness of tasks is also in line with theories of innovation arguing that a single company innovating by itself can increase its innovation capacity (idea generation, as well as conversion of ideas to marketplace) through a network of strategic partnerships (Hansen and Birkinshaw, 2007). This type of motivators for interorganizational relationships is aligned with the market view of innovation being an external driver of innovation in companies focusing on product quality and (Paladino, 2007). Observing further the motivator customer value of interorganizational relationships through the prism of reinforcing interorganizational relationships, it seems to also connect well with innovation theories arguing that social network relationships seem to be of a vast importance to organizations, as through social capital information is exchanged and knowledge transfer between organizations is fostered (Borgatti and Foster, 2003).

¹ RBV – Resource Based View

It should be noted that a large number of innovation theories (Zhang et al., 2010; Andersen and Drejer, 2008; Murray and O'Mahony, 2007; Shu-hsien et al., 2007) argue that one of the most important drives influencing innovation performance is knowledge management (creation, absorption, transfer, accumulation and recombination of knowledge). Most important for innovation is new knowledge which is typically created by recombining and reusing the existing knowledge (Kogut and Zander, 1992). Companies who source knowledge from outside of the boundaries of their organizations are more likely to have access to a broader knowledgebase, rather than if the source of knowledge is only within the organization's own boundaries (Easterby-Smith et al., 2008; Drejer, 2008). Similarly, cooperation between a company and a number of partners can be observed as a distributed organization; hence such organization is having access to a larger pool of knowledge (Dejer, 2008). Having access to a broader knowledgebase through external relationships is very important for companies seeking to increase innovation performance – as through collaboration with external partners companies are more likely to achieve a better new knowledge creation and hence innovation performance than if they were to innovate just by themselves (Fallah and Lechler, 2008; Mrinalini and Nath, 2008; Goyal and Pitt, 2007). Similarly, innovation performance of an individual organization seems to be in a close relationship with IORs - collaborative relationships amongst organizations are found to positively influence innovation performance (Palmatier, R. et al., 2007; Davis and Eisenhardt, 2007; Faems et al., 2005). In transferring knowledge amongst organizations, it is important to recognize knowledge absorption capacity as an ability to integrate new knowledge from its network of external organizations. The knowledge absorption capacity of an organization also dictates the speed of integration of new knowledge (Soosay and Hyland, 2008; Shu-hsien et al., 2007). On the other hand organization's knowledge desorption capacity - an ability of an organization to transfer its own knowledge to external partners (i.e. outbound technology transfer) is also significant attribute of interorganizational activity in innovation networks (Lichtenthaler et al., 2010).

We also need to take into a consideration that organizations operate within the wider social circumstances, and that organization's prominent positioning within the social environment affects is ability to transfer knowledge (Parmigiani and Riviera-Santos, 2011). Research also indicates that partner type in a network of companies influences outcome of innovation performance (Changsu and Jong-Hun, 2010; Kang and Kang, 2010; Faems *et al.*, 2005). It is interesting to note that researchers (Changsu and Jong-Hun, 2010; Kang and Kang, 2010; Bergman and Maier, 2009) argue that if the partner type is a university collaborating on R&D activities with a company, the innovation performance will be significantly higher as compared to other partner types.

It should also be noted that interorganizational relationships between companies can be stronger or weaker; companies needs both types of relationships since stronger relationships provide deeper exchanges fostering organizational learning and innovation, while weaker relationships can be established with wider range of partners, providing the company with a range of different knowledge and resources (van Wijk *et al.* 2008). This is in line with the theories of innovation indicating that innovation outcome, speed and effectiveness is influenced by the strength of interorganizational relationships, however also by the number of relationship and a density of the partner network. Researchers argue that larger the partner network size and strength of links amongst companies is such organization will have a stronger performance (Oviatt and McDougall, 2005).

Researchers (Huggins and Johnston, 2009; Valliere and Peterson, 2009; Simmie, 2004) argue that relationships between companies are no longer of a local geographic character only, rather innovation driven relationships amongst organizations are becoming increasingly of a regional and international character – indicating a more complex interorganizational relationships (i.e. innovation networks) across national boundaries. Growth in developing countries seems to be fuelled by imitation of globally novel innovations (adopted innovations that are novel only at the industry or national level), investment in knowledge attainment and R&D activities, rather than through producing globally novel innovations. In addition, intensity of research, educational level and capacity to absorb foreign technology stimulate innovation activities and seem to have a profound effect on productivity growth of a country (Madsen *et al.*, 2010). Furthermore, growth in developed countries² is in part fuelled by the growth in developing countries typically through investing in developing countries and reaping benefits from a growth that is typically faster than that of a developed world (Nasrolahi et al., 2010; Madsen et al., 2010; Wang, 2010). On the other hand, growth in developing countries can be propelled with knowledge transfer from developed countries and investments in R&D (Madsen et al., 2010). In technology sector this is typically realized through companies in developed world investing in R&D in foreign countries as a substitute for their internal R&D efforts (such as for example outsourcing). Some of the reasons for that is the cost-effective of R&D in developing countries and acquisition of highly skilled human resources which seems to be in a large demand in the developed world (Wang, 2010). Effectiveness in knowledge transfer from developed to developing countries in conjunction with knowledge levels is also influenced by developing country's investment in R&D - as these resources represent vehicles of knowledge absorptive capacity (Madsen et al., 2010). It should be noted that "it is practically impossible to do things identically" indicating that every change (e.g. transfer of technology from developed world to SEE) is an innovation (Hansen and Wakonen, 1997). As such, it is plausible to believe that innovations that are transferred from developed to developing countries (e.g. SEE) in order to be adopted need to undergo a change in order to be implemented (adopted) into the environment of the host country, and therefore can be classified as innovations novel for the particular host country.

² Developed countries are classified as all countries with more than annual USD \$20,000 per capita, whereas undeveloped countries are classified as all countries with less than \$20,000 per capita (Valliere and Peterson, 2009)

3. Methodology

Desk research was used to build a theoretical framework as a foundation of the study. In conducting the desk research and building a theoretical framework as a basis of the study and constructions of the research instrument online database resources were used (EBSCO, Business Source Complete, The British Library Theses Service & Electronic Thesis Online Service, Google Scholar). Strategy pursued in conducting a systematic literature review has consisted of the following process followed:

- Researched 20 full text PhD studies on innovation with the filter from 2000-2011 in order to understand the areas of research in innovation and directions other researchers were pursuing.
- In order to understand the most current literature on the topic, the first article search was performed very broadly on a single unrestricted keyword "innovation", however with a four year filter (2009-2012) in order to identify all articles related to innovation within the most recent years. This process lead to downloading almost 200 journal articles for further study and bibliographical review.
- Further the process has included grouping and compiling the set of publications in order to understand the areas of innovation covered broadly by the current academic literature. One of the challenges was to filter out articles not related to innovation management in particular, as the term innovation was also used to describe creativity as well.
- Following the initial process of literature review, a more targeted research of journal articles was performed with the aim to focus on articles that were of a particular interest for this research topic, expanding the year range from year 2000 and onwards. The particular keyword terms searched are:
 - Innovation theory
 - Systems of innovation
 - Knowledge and innovation, knowledge spillovers
 - Absorption capacity and knowledge
 - Technology transfer, knowledge transfer and innovation
 - Clusters, innovation clusters, technology clusters
 - Network innovation., network path analysis
 - Open Innovation
 - Interorganizational relationships theory
 - Interorganizational relationships, IOR and innovation
 - Organizational theory, organizations economic theory
 - RBV and innovation, RBV and IOR
 - Developed and undeveloped countries and innovation

Following this process further selection of academic articles was performed based on the number of citations – articles with the highest number of citations were selected for further review. It is interesting to note that Saha *et al.* (2003) argue that number of citations represent a de facto standard of quality of academic articles and contribution to the knowledge.

Broad scope view on innovation was focused on specific topics of interest further in the literature review through research funnel methodology (Berthon *et al.*, 2003) – starting wide and narrowing topics down to our focus research area. According to Berthon *et al.* (2003) this methodology and approach will help us observe and identify anomalies between observations and assumptions, including a more clear identification of articles that challenge current assumptions, hence provide more insightful research thinking.

4. Analysis

In reviewing current theories of innovation and organizational theories it can be noted that the academic literature seems to be lacking theoretical frameworks connecting innovation and organizational theories, as such representing the area of research opportunity. Supporting this view is Crossan and Apaydin (2010) arguing that there exists no consolidation of academic literature on innovation and that there exist gaps in understanding relationships between individual innovation facets present in the literature. Further, Parmigiani and Rivera-Santos (2011) argue that academic research on interorganizational relationships typically observes IORs from a single point of view - either as a form of interorganizational relationship, as a process between interorganizational relationships, and outcome of interorganizational relationships – hence indicating that a more integrated and comprehensive view of IORs is required. Santos and Eisenhardt (2005) suggest that in order to be able to adequately observe interorganizational relationships we cannot look separately into the economic and sociological factors motivating the establishment of interorganizational relationships, arguing that both approaches have to be taken into a consideration; therefore suggesting that a multi-theoretical approach is required. Furthermore, Crossan and Apaydin (2010) also argue that innovation should not be observed only through a single prism of either a process or an outcome, as the two are dependent upon each other and should be observed jointly. It also seems that academic body of knowledge is not enriched with specific coverage on interorganizational relationships between developed world and SEE^3 and its influence to innovation outcome, and especially with focus on technology service companies. From this analysis it seems that current research is indicating gaps in the academic literature representing research opportunities. The analysis also indicates a need for multi-theoretical approach in consolidating both organizational and innovation theories. Combining organizational and innovation theories therefore seems to indicate a very novel direction in which the topic can be researched further. As this area of academic knowledge is not well researched, this would seem to indicate that due to the lack of theoretical support of influence of IORs to innovation performance, perhaps the best methodology for conducting the further research would be exploratory.

In combining organizational and innovation theories, research indicates there are matching motivators of existence of IORs and drivers of innovation in the two main directions. One is viewing motivators for creation of interorganizational relationships

³ SEE – South East Europe

when a specific activity can be conducted in a more efficient way (Parmigiani and Rivera-Santos, 2011) and is matched with the RBV drivers of innovation (Paladino, 2007). The RBV innovation theories support a view that relationships in which a group of companies innovates jointly is more suited for manufacturing sector seeking to improve a new product development (Andersen and Drejer, 2008; Paladino, 2007). The other view is that organizations form IORs in order to strengthen their core competencies through seeking partnerships with others companies and entities (Parmigiani and Rivera-Santos, 2011). This approach suggests that a single company is innovating and leveraging knowledge from its partner network. In this area knowledge transfer amongst organizations and social networks seem to be predominant in influencing innovation outcome. Innovation theories support that that social relationships and tacit knowledge transfer is more aligned innovation strategy fostering innovation in service based companies (Murray and O'Mahony, 2007; Shuhsien et al., 2007). As this driver of existence of IORs is more aligned with the service sector companies focusing on higher customer value and quality of service, it seems that researching in this direction is more aligned with our target group of observing technology service companies - identified as industry with the most of the innovation activity (Fallah and Lechler, 2008). Therefore analysis of the literature seems to indicate that observing interorganizational relationships from the view of knowledge transfer and fostering interpersonal \ social relationships is more in line with our research group of technology service companies providing further focus of the research.

Expanding our view further on influence of interorganizational relationships to innovation outcome, the literature seems to indicate that knowledge management, and knowledge transfer is one of the most important factors of a relationship influencing innovation outcome. Companies who have access to a larger body of knowledge, through interorganizational relationships, can be more successful in innovation compared to companies not leveraging knowledge transfer and generating new knowledge just by themselves (Fallah and Lechler, 2008; Mrinalini and Nath, 2008; Goyal and Pitt, 2007). This is also supported by a large number of innovation theories (Zhang et al., 2010; Andersen and Drejer, 2008; Murray and O'Mahony, 2007; Shuhsien et al., 2007) arguing that innovation performance is directly related to knowledge management and that new knowledge is generated through recombining and reusing the existing knowledge. On the other hand, this relationship is also supported by a number of organizational theories arguing that several forms of partnerships types contribute to learning, knowledge creation and knowledge dissemination. Researchers (Parmigiani and Rivera-Santos, 2011; Lavie et al., 2010) argue that alliances amongst organizations contribute to better learning innovation; cross-sector partnerships contribute to new knowledge creation and organizational networks contribute to a broader knowledge flow that can contribute to innovation. Analysis on combining both innovation and organization theories indicates there is a strong support that interorganizational knowledge transfer has a strong influence to innovation outcome, therefore indicating further focus of the research on this specific aspect of IORs.

However, transferring knowledge amongst organizations is not without challenges for both the host (organization initiating knowledge transfer) and recipient (organization receiving knowledge). It seems that research indicates that on the recipient end the knowledge absorption capacity limits the volume and speed of knowledge absorption (Soosay and Hyland, 2008; Shu-hsien et al., 2007), whereas on the host end in order for the knowledge transfer to occur in the first place there also has to be knowledge desorption capacity – willingness and capability of the host to transmit the knowledge to the recipient (Lichtenthaler et al., 2010). Furthermore, it also seems that organization's position within the knowledge network and partnership types also have an influence on the success and nature of the knowledge transfer (Parmigiani and Riviera-Santos, 2011). In this respect it is interesting to note that the research (Changsu and Jong-Hun, 2010; Kang and Kang, 2010) indicates that if companies collaborate on an R&D project with a research university, there is a likelihood of a much better innovation performance compared to other types of partners. On the other hand not only partner type, but also strength of a relationship influences the success of the interorganizational knowledge transfer (Oviatt and McDougall, 2005). It seems that the strength of relationships is perhaps the most appropriate to observe through the aspect of social relationships as they seem to play a very important role influencing the knowledge transfer, especially tacit knowledge transfer (Borgatti and Foster, 2003). This is important from the standpoint of innovation theories arguing that innovation is strongly driven by tacit knowledge transfer (Madsen et al., 2010). This is also in agreement with organizational theories arguing that knowledge flow amongst organizations seems to be influenced, amongst others, through social networks (Parmigiani and Riviera Santos, 2011). In order to stimulate such knowledge exchange, the research indicates that social relationships play the most important role, as through social interaction experiences (tacit knowledge) are more closely relayed. Traditionally academic literature would observe social relationships amongst companies as restricted to a specific geographic area (e.g. technology clusters), however current research indicates that these relationships are no longer of a local character and that innovation driven relationships are increasingly having a strong cross-border character. This is supported through development of information technology supporting social networks – such networks can server as a main enabler of social interactions amongst companies (Borgatti and Foster, 2003). Analysis and integration of the current research between innovation and organization theories seems to indicate that there are number of individual determinants of successful interorganizational transfer, however moving beyond the transfer itself such knowledge transferred has also to be applied in an appropriate way in order to influence successful innovation outcome. As such, it can be argued that if determinants of successful knowledge transfer, coupled with innovation theories on determinants of successful innovation can provide a framework for measuring influence of interorganizational knowledge transfer to innovation outcome – as a topic of the further synthesis and research efforts.

The academic literature seems to indicate there are several aspects specific to a successful knowledge transfer from the developed world to developing countries. Absorption capacity of foreign knowledge along with the intensity of research, and educational level has a significant effect on innovation performance (Madsen *et al.*,

2010). Effectiveness in knowledge transfer from developed to developing countries in conjunction with knowledge levels is also influenced by developing country's investment in R&D – as these resources represent vehicles of knowledge absorptive capacity. It should also be noted that economies in developing world are fuelled predominantly through imitation of the existing global innovation (Madsen *et al.*, 2010). Introducing an existing innovation from a developed into a developing country will indeed be an innovation novel to the receiving country – as supported by Hansen and Wakonen (1997) arguing that every change is an innovation by itself. It is therefore plausible to believe that innovations that are transferred from developed to developing countries in order to be adopted need to undergo a change in order to be implemented (adopted) into the environment of the host country, and therefore can be classified as innovations novel for the particular host country. Analysis points out that the most commonly seen examples of innovation in SEE can perhaps be found in innovations novel to the region or specific to a country only, and not necessarily globally novel innovations – presenting further field research opportunities.

5. Conclusions

Companies no longer have a choice whether they should innovate or not – noninnovative companies simply cannot sustain the forces of the global competition. External organizational relationships are very important for a company, as no single organization can survive for a prolonged period of time on its own in today's global environment. Innovation performance of an individual organization seems to be in a close relationship with interorganizational interactions; collaborative relationships amongst organizations are found to positively influence innovation performance. Furthermore, technology service companies are found to have the most of the industrial innovation activity. These are the reasons as why interorganizational relationships amongst technology service companies and their influence to innovation outcome were chosen as a topic of this study.

Review of the academic literature indicates that current research lacks theoretical frameworks integrating theories of innovation and interorganizational relationships, as such representing the area of research opportunity. It also seems that academic body of knowledge is not enriched with specific coverage on interorganizational interactions between developed world and SEE especially in the technology sector and in relationship to innovation. Findings indicate that it seems that one of the most important links of interorganizational relationships influencing innovation performance is interorganizational knowledge transfer. This is supported by a large body of innovation theories arguing that knowledge management, accumulation, reuse and recombination of knowledge influences generation of new knowledge and is attributing to a successful innovation performance. These innovation theories are in line with organization theories arguing that new knowledge creation and a broader knowledge flow are fostered through alliances amongst organizations. The knowledge transfer amongst organization seems to be driven by a number of determinants – knowledge absorption \ distortion capacity, type of a partner, position within a network, strength of the relationships, social networks and others – supported by both innovation and organization theories. Furthermore it should be noted that tacit knowledge transfer seems to be a relevant influencer of knowledge recombination and new knowledge creating leading to innovation. In order to foster tacit knowledge transfer reinforcing the strength of social networks and social relationships is necessary. These types of relationships are more aligned with service companies that seem to be in line with our target group of researching technology service companies. Further applying these relationships to transferring knowledge from developed to developing countries; a vehicle of knowledge absorption capacity is based on resources – investment in R&D and knowledge levels of the recipient organization. It should also be noted that innovation in developing countries should be observed at the level of imitation, and not necessarily creation of globally novel innovations. This is supported through views that what is transferred to the recipient country has to be adapted to the local circumstances, and such adaptation creates a novelty – innovation.

From the desk research presented in this paper theoretical gaps between organization and innovation theories are evident, along with their potential application to SEE – representing research opportunities. The analysis of the desk research indicates that perhaps the most logical further focus of the study should be on exploring determinants of interorganizational knowledge transfer and its influence to innovation outcome.

References

- Andersen, P. and Drejer, I. (2008) Systemic innovation in a distributed network: the case of Danish wind turbines, 1972-2007. *Strategic Organization*, 6(1), p.13-46
- Bergman, E. and Maier, G. (2009) Network Central: Regional Positioning for Innovative Advantage. *Annals of Regional Science*, 43(3), p.615-644
- Berthon, P. et al. (2003) Through the Paradigm Funnel: A Conceptual Tool for Literature Analysis. *Marketing Education Review*, 13(2), p.55-66
- Borgatti, S. P., and Foster, P. C. (2003) The network paradigm in organizational research: A review and typology. *Journal of Management*, (29), p.991-1013
- Burgess, S. (2011) Measuring Financial Sector Output and Its Contribution to UK GDP. *Bank of England Quarterly Bulletin*, 51(3), p.234-246
- Changsu, K. and Jong-Hun, P. (2010) The Global Research-and-Development Network and its Effect on Innovation. *Journal of International Marketing*, 18(4), p.43-57
- Clayton, C. et al. (2001) The Great Disruption. Foreign Affairs, 80(2), p.80-95
- Crossan, M. and Apaydin, M. (2010) A Multi-Dimensional Framework of Organizational Innovation: A Systematic Review of the Literature. *Journal of Management Studies*, 47(6), p.1154-1191
- Davis, J. and Eisenhardt, K. (2007) *Rotating Leadership and Symbiotic Organization: Relationship Processes in the Context of Collaborative Innovation.* Sloan School of Management, MIT, Cambridge.
- Faems, D. et al. (2005) Interorganizational Collaboration and Innovation: Toward a Portfolio Approach. Journal of Product Innovation Management, 22(3), p.238-250

- Fallah, M. and Lechler, T. (2008) Global innovation performance: Strategic challenges for multinational corporations. *Journal of Engineering & Technology Management*, 25(1), p.58-74
- Gilbert, C. (2003) The Disruption Opportunity. *MIT Sloan Management Review*, 44(4), p.27-32
- Goyal, S. and Pitt, M. (2007) Determining the role of innovation management in facilities management. *Facilities*, 25(10), p.48-60
- Hansen, M. and Birkinshaw, J. (2007) The Innovation Value Chain. Harvard Business Review, 85(6), p.121-130
- Hansen, S. O. and Wakonen, J. (1997) Technology Transfer, a Public Utility?. *Liiketaloudellinen Aikakauskirija*, 46(3), p. 349-61
- Huggins, R. and Johnston, A. (2009) Knowledge Networks in an Uncompetitive Region: SME Innovation and Growth. *Growth and Change*, 40(2), p. 227-529
- Kang, K. H. and Kang, J. (2010) Does Partner Type Matter in R&D Collaboration for Product Innovation?. *Technology Analysis and Strategic Management*, 22(8), p.945-59
- Kogut, B. and Zander, U. (1992) Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, 3(3), p.383–397
- Lavie, D. *et al.* (2010) Exploration and exploitation within and across organizations. *Academy of Management Annals*, 4(1), p.109-155
- Leiponen, A. (2000) Competencies, Innovation and Profitability of Firms. *Economics* of Innovation and New Technology, 9(1), p.1-24
- Lichtenthaler *et al.* (2010) Technology Transfer across Organizational Boundaries: Absorptive Capacity and Desorptive Capacity. *California Management Review*, 53(1), p.154-170
- Madsen, J. *at al.* (2010) Catching up to the technology frontier: the dichotomy between innovation and imitation. *Canadian Journal of Economics*, 43(4), p.1389-1411
- McDougall, P. and Oviatt, B. (1996) New Venture Internationalization, Strategic Change, and Performance: a Follow-up Study. *Journal of Business Venturing*, 11(1), p.23-42
- Moore, J. (1993) Predators and prey: A new ecology of competition. *Harvard Business Review*, 71(3), p.75–86
- Mrinalini, N. and Nath, P. (2008) Knowledge Management in Research and Technology Organizations in a Globalized Era. *Perspectives on Global Development & Technology*, 7(1), p.37-54
- Murray, F. and O'Mahony, S. (2007) Exploring the Foundations of Cumulative Innovation: Implications for Organization Science. Organization Science, 18(6), p.1006-1021
- Nasrolahi, K. *et al.* (2010) Impact of Technology Transfer on Research and Development Activities in Developing Countries During The Period 1999-2008. *Interdisciplinary Journal of Contemporary Research in Business*, 3(4), p.773-783
- Paladino, A. (2007) Investigating the Drivers of Innovation and New Product Success: A Comparison of Strategic Orientations. *Journal of Product Innovation Management*, 24(6), p.534-553

- Palmatier, R. et al. (2007) A Comparative Longitudinal Analysis of Theoretical Perspectives of Interorganizational Relationship Performance. Journal of Marketing, 71(4), p.172-194
- Parmigiani, A. and Rivera-Santos, M. (2011) Clearing a Path Through the Forest: A Meta-Review of Interorganizational Relationships. *Journal of Management*, 37(4), p.1108-1136
- Prajogo, D. (2006) The Relationship Between Innovation and Business Performance A Comparative Study between Manufacturing and Service Firms. *Knowledge & Process Management*, 13(3), p.218-225
- Saha, S. et al. (2003). Impact factor: a valid measure of journal quality. Journal of the Medical Librarian Association, 91(1), p.42–60.
- Santos, F. and Eisenhardt, K. (2005) Organizational Boundaries and Theories of Organization. *Organization Science*, 16(5), p.491-508
- Selden, L. and MacMillan, I. (2006) Manage Customer-Centric Innovation Systematically. Harvard Business Review, 84(4), p.108-116
- Sen, C. (2011) FDI in the Service Sector-Propagator of Growth for India?. *Theoretical and Applied Economics*, 18(6), p.141-56
- Shu-hsien, L. (2007) Knowledge sharing, absorptive capacity, and innovation capability: an empirical study of Taiwan's knowledge-intensive industries. *Journal of Information Science*, 33(3), p.340-359
- Simmie, J. (2004) Innovation and Clustering in the Globalised International Economy. *Urban Studies*, 41(5), p. 1095-1112
- Soosay, C. and Hyland, P. (2008) Exploration and exploitation: the interplay between knowledge and continuous innovation. *International Journal of Technology Management*, 42(1), p.20-35
- Tether, B. (2005) Do Services Innovate (Differently)? Insights from the European Innobarometer Survey. *Industry & Innovation*, 12(2), p.153-184
- Valliere, D. and Peterson, R. (2009) Entrepreneurship and Economic Growth: Evidence from Emerging and Developed Countries. *Entrepreneurship and Regional Development*, 21(5), p.459-480
- van Wijk, R., *et al.* (2008) Inter and intra-organizational knowledge transfer: A metaanalytic review and assessment of its antecedents and consequences. *Journal of Management Studies*, 45(1), p.830-853
- Zhang, H. et al. (2010) Managing Knowledge for Innovation: The Role of Cooperation, Competition, and Alliance Nationality. Journal of International Marketing, 18(4), p.74-94

Strategic Planning Practices in the State Owned Banks The Indonesian Case

Mohammad Sihab Ridwan and John Marti

Southampton Management School, Faculty of Business and Law, University of Southampton, England, United Kingdom, Email: msr1e09@soton.ac.uk

Abstract. Although the importance of strategic planning practices has been recognized in academic and practitioner literature, for the Indonesian case no studies in the area of strategic planning practices have been done for both the government and private sector. To the best of the researcher's knowledge, it is the first study trying to analyze the issue under the Indonesian case. A qualitative research method (Merriam, 2002) was used as the framework for this study. It will investigate and write down the strategic planning practices at the state owned banks in Indonesia to study strategic planning practices. Founded on the previous study, this study developed a research model by combining key constructs of strategic management with those of organizational context. A multiple-case study (Yin, 2009) was adopted as a research strategy to comprehend a real-life phenomenon in depth. Investigation on business companies has often assumed the form of case studies (Yin, 2003). The field data were analyzed using the theoretical framework of the model. The six generic steps as suggested by Creswell (2009) ware used to explain the process of data analysis of this investigation. Thematic analysis (Miles and Huberman, 1994) will be employed to analysis the data. The preliminary findings provide researcher an comprehending concerning with stages, processes of planning practices, styles of decision making, and corporate performance at each of three state owned banks.

Keywords: *Strategic Planning, Organizational Context, Performance*

1 Introduction

Almost fifty years strategic planning has become a standard part of management thinking and practice in the business world and more than twenty five years it has become the standard practice of large numbers of public and non-profit organization (Bryson, 2006). The first and perhaps most obvious benefit of strategic planning is the promotion of strategic thinking, acting, and learning (Van der Heijden, 1996). However, there is a gap between planning process including implementation and the extent to which organizational context supports or distracts from planning process (Rowley & Sherman, 2002). Other factors will impact on the relationship between strategic planning and performance (Rudd et.al, 2008). Strategic planning often fails due to problems or barriers encountered at the implementation stage (O"Regan and Ghobadian, 2002). Mixed evidence about the relationship between strategic planning and organizational performance makes the debate about its effectiveness as a tool of strategic management an on-going one (Obeng and Ugboro, 2008). The linkage between strategic planning, organizational context, and firm performance needs analysis to get a better understanding how strategic planning is applied in practice and will improve organizational performance.

1.1 Research Objectives

• To explore and document of the dynamics of planning practices (changes) at the high-performing banking industry in the state owned banks in Indonesian context.

• To investigate the role of organizational context to facilitate strategic planning practices and how they led to the bank's successful performance such as growth in sales and profits.

• To extend the knowledge of strategic planning practices and intends to fill the gap in the literature.

1.2 Research Questions

• How do managers in the state owned banks develop and implement strategic planning in practice?

• To what extent is the relationship between strategic planning practices and firm performance?

• To what extent will culture and decision approaches contribute to building an organizational context to facilitate successful strategic planning?

2. The Literature Review

2.1 General description

Strategic planning is inseparably connected into the entire fabric of management; it is not something split and distinct from the process of management (Steiner, 1979). Strategic planning is a function and concern of entirely managers at all levels in an organization. To take a broad view, according to Steiner, there are two

types of management. That which is completed at the highest of an organizational structure is strategic management; everything else is operational management. "Strategic planning is a backbone support to strategic management" (P.4), although, of course, strategic planning is not the wholeness of strategic management; it is a main process in the running of strategic management. Strategic planning is central to helping managers satisfy their strategic management everyday jobs. Just as strategic management is absolutely concerned with operational management so as strategic planning linked with operations. But as with strategic management, the focus and highlighting of strategic planning is on strategy more than operations.

Further, Steiner presented that there are two primarily different ways for a manager to formulate strategic plans for the time ahead. The first, the intuitiveanticipatory approach. It has several main features; mostly it is made in the brain of one person and may not result in a written set of plans. Commonly it has a reasonably short time horizon and reaction time and "it is based upon the past experience, the gut feel, the judgment, and the reflective thinking of a manager, " (P. 9). However, how many times are intuitive truthful in the judgements?

In contrast, the formal planning system is structured and developed on the basis of set procedures. It is obvious in the sense that people understand what is going on. Regularly, manuals of operation are prepared to describe who is going to do what and when and what will occur with the information. It is based on research and engages the participation of many people. Support for the decision making in the process is regularly documented and the result of whole effort is a written set of plans.

Both of them can and should complement one another. A formal system can and should facilitate managers to sharpen their intuitive-anticipatory ideas into the planning process. At the very least, a formal system can and should give managers more time for reflective thinking. However, as Steiner has presented that formal planning system cannot be really effective unless managers at all levels shoot up their judgements and intuition into the planning process. Nor, conversely, will formal planning be effective if top managers refuse it in support of their own intuition. As Mintzberg (1994) said that to be effective, any organization has to combine analysis and intuition in its strategy making as well as other processes. In addition, Steiner explained that managers, in reality, do stick to diverse thought processes in decision making. The blueprint of a formal planning system must understand and indicate these divergences if the system is to perform successfully.

2.2 Some definitions of strategic planning

One of the seminal works in strategic planning, Anthony's work (1965:24), defined strategic planning as "the process of deciding on objectives of the organization, on changes in these objectives, on the resources used to attain these objectives, and on the policies that are to govern the acquisition, use, and disposition of these resources." Here, strategic planning is a process to formulate strategic plans and policies transforming the character or direction of the organization. Further, Anthony explained that in an industrial firm this process comprises planning influencing the objectives of the firms, all types of policies, market and distribution channels, the acquisition and disposition of main facilities, new permanent capital

sources, divisions, subsidiaries organization structure, research and development of new product lines, and so on.

Another seminal work is Steiner's work (1979) that described strategic planning from several points of view, one of them is strategic planning is a structure. The concept of a structure is reflected in this definition: "strategic planning is the systematic and more or less formalized effort of a company to establish basic company purposes, objectives, policies, and strategies and to develop detailed plans to implement policies and strategies to achieve objectives and basic company purposes" (P.15).

The other seminal work is Lorange's work (1980) describing strategic planning as "the process of deciding on objectives of the organization, on changes in these objectives, on the resources used to attain these objectives, and on the policies that are to govern the acquisition, use, and disposition of these resources" (P.24). Further, Lorange (2010) presented that the aim of strategic planning is obvious: to be a management tool to aid in strategic decision making process of an organization. If an activity that runs under the strategic planning label and it does not help in strategic decision making of the organization it is not strategic planning even though there are alot of right planning components involved in the activity, such as details five-year plan documents. Another definition by Bryson (1995:4-5) described strategic planning as "a diciplined effort to produce fundamental decisions and actions that shape and guide what an organization is, what it does, and why it does it."

The last and perhaps the most comprehensive and descriptive definition as also said by Ansoff (1965, 1968) was provided by Peter Drucker (1959:240). He defined strategic planning (long range planning) as:

"The continues process of making present entrepreneurial (risk taking) decisions systematically and with the best possible knowledge of their futurity, organizing systematically the efforts needed to carry out these decisions, and measuring the results of these decisions against the expectations through organized, systematic feedback." Ansoff pointed out the Drucker's definition in three parts:

1. Making present decisions systematically

2. Organizing programmes for their implementation

3. Measuring actual performance of these decisions (the programmes).

The linkage of Drucker's definition to Ansoff's work (Corporate strategy) is their entire concern with "making present entrepreneurial decisions" (Ansoff: P.185). To broaden his results to Drucker's definition of strategic planning, Ansoff creates a document called a strategic budget which will organize systematically the attempt for execution the decision. In addition, Drucker presented that it is easier to define longrange planning by what it is not rather than by what it is, namely:

1. It is not "forecasting". It is not engineering the future and any effort to do so is thoughtless: human beings can neither foresee nor control the future.

2. It does not cope with future decision. It copes with the futurity of present decision. Decisions occur only in the present. The question that meets the long-range planner is not what we should do tomorrow. It is what we have to do today to be prepared for an ambiguous tomorrow

3. Long-range planning is not an effort to exclude risk. It is not even an effort to minimize risk. Indeed any such effort can only run to illogical and unlimited risk and to certain failure.

2.3 A brief history of strategic planning

Considered in the setting of human thought and behaviour, strategic planning is definitely not new (Dooris, Kelley, and Trainer (2002). To the contrary, since strategic planning represents vital features of Homo sapiens, it is by definition as old as humankind. Conversely, when one views strategic planning as a well-thought-out management discipline and practice, it is barely out of its infancy. Similarly, Lorange and Vancil (1977) presented that strategic planning as the form of human being mental activity is not new; what was new was the attempt by managers in large firms to reinforce strategic planning in early 1960s.

In addition, Barker and Smith (1997) described that the roots of strategy and planning concepts can be traced to military applications which the literature on strategy dates from classic writings of Sun-tzue (Art of war, 6th Century BC) and later, writings of Carl von Clausewitz (on war) in 19th Century. The word strategic suggests a close relationship with strategy as used in military parlance in which military leaders and researchers have done much thinking about strategic principles (Anthony, 1965). Business management should be able to take benefit from what the military leaders and researchers have studied and published. Further, Anthony depicted that of the fourteen well-known planning principles set forth by Harold Koontz in 1964, the nine planning principles have relationship to the military principles.

Strategy can be understood as "the what is to be done and planning as the how to do it" (Barker and Smith, 1997:289). Drucker (1974) noted as also quoted by Barker and Smith that in the early 1920s for the first time Du Pont developed systematic approaches to business objectives, to business strategy, and to strategic planning. Impetus was given to strategic planning in the post-World war II period in which many who had involved in the US armed forces was back to civilion life and applied the concepts of strategy and planning to firm and industry (Barker and Smith, 1997). Long range planning evolved and as argued by Ansoff (1990:247) that in the 1950s "long range planning was the firm"s response to the pressures of rapid growth, size, and complexity" which firms were no longer likely to rely on budgeting in facing their future competitive challenges, growth and expansion needs.

Strategic planning arose as a distinct methodology for a time between the 1950s and the 1970s (Dooris, Kelley, and Trainer, 2002). Lorange and Vancil (1977) wrote that In early 1960s, managers in large firms attempted to formalize strategic planning and to focus it on influencing the strategic directions of their organizations. In that 1960s, business policy and strategic planning have played increasingly important roles in developing formal alternatives to increase firm performance, and more recently there have been a proliferation of major works (Shrader, Taylor, Dalton, 1984). Steiner (1979) asserted that strategic planning with its modern design characteristics was first introduced in profit oriented companies in the mid-1950s by large companies under the rubric of long-range planning system. Since then, according to Steiner, formal strategic planning has matured and all large firms

throughout the world virtually have adopted strategic planning system and a number of smaller firms are adopting the example of the larger firms. Mintzberg (1994) noted that strategic planning arrived on the scene in the mid-1960s when corporate leaders embraced it as the one best way to device and implement strategies that would enhance the competitiveness of each business unit.

The 1960s, according to Steiner, was a period of steady economic growth and general prosperity in United States and during this period there were many attractive opportunities for growth and corporate executives realized that they had to choose sometimes to diversify through acquisition and to expand to international market, and so on. These strategic moves raise the managerial complexity of large enterprises, and new management tools were obviously needed to help corporate executives overcome an increasing complexity of strategic decisions. Formal long range planning almost looks like a godsend to these corporate executives and its main virtue was that "it focused on the right set of issues" (Steiner, 1979:P.vii). During the five years of 1968-1972, the development of strategic planning systems established very fast. Hundreds of executives and scores of academicians were devoted to learn about the design of strategic planning systems. Others characteristic the emergence of strategic planning to the turbulent environment of 1970s when, with the energy crisis and other unforeseen events, organizations scurried to find a more appropriate planning system (Rosenberg and Schewe, 1985). Many would claim that investigating for the birthstone of strategic planning is chimerical since planning is an evolutionary course (Dooris, Kelley, Trainer, 2002). Certain dating stones can be traced, but strategic planning owns no single event of origin. What is obvious, however, is that the latest some decades have been a prosperous period for strategic planning- "a development in which higher education has shared" (P.6). The concept of strategic planning has progressed over time and from its early stages in companies has spread to government and public sector as well as higher education (Barker and Smith, 1997).

2.4 Dimensions of Strategic Planning Process (Strategic Planning Practices)

The author extends the theoretical review of 17 Dimensions of Strategic Planning Process identified by Wegner (2006) as follows:

1. Awareness of financial constraints (Lorange & Vancil, 1977; Steiner, 1979; Barker & Smith, 1997; Keller, 1983; Kotler&Murphy, 1983; Bryson, 2004).

2. CEO or president is involved and a champion for planning (Barker &Smith, 1997; Keller, 1983; Kotler&Murphy, 1983; Steiner, 1979; Lorange and Vancil, 1977; Rowley &Scherman, 2002; Bryson, 2004; Heines, 1995).

3. Clear and measurable objectives (Ansoff, 1965; Lorange & vancil, 1977; Steiner, 1979, Heines, 1995; Bryson, 2004:)

4. Commitment to take action even if only small wins are achieved at first (Keller,1983;Meredith,1993;Steiner,1979;Bryson,2004;Heines,1995.

5. Communication lines established and open (Barker &Smith, 1997; Kotler&Murphy, 1983; Meredith, 1993; Steiner, 1979).

6. Environmental scanning (Ansoff, 1965, Lorange & Vancil, 1977; Barker & Smith, 1997; Keller, 1983; Kotler&Murphy, 1983; Steiner, 1979, Bryson, 2004;

Heines, 1995; Hunger & Wheelen, 2003; Ardekani & Haug, 1997; Dooris , keller, Trainer, 2002)..

7. Feedback and evaluation (Ansoff, 1965; Antony, 1966; Lorange and Vancil, 1977; Kukalis, 1991; Heines, 1995; Barker &Smith, 1997; Steiner, 1979; Bryson, 2004; Wheelen&Hunger, 2003).

8. Flexibility to adjust planning process and goals after process is complete (Poland&Arns, 1978; Lorange & vancil, 1977; Steiner, 1979; Liedka, 2000).

9. Hierarchical planning (planning occurs at each institutional level), (Lorange & Vancil, 1977; Barker &Smith, 1997; Keller, 1983; Kotler&Murphy, 1983; Steiner, 1979; Bryson, 2004).

10. Information flow and communications (Lorange and Vancil, 1977; Barker &Smith,1997;Keller,1983;Steiner,1979,Liedka,2000,Bryson,2004).

11. Linking of goals to budgets (Ansoff, 1965; Lorange and Vancil, 1977; Barker &Smith, 1997; Keller, 1983; Kotler&Murphy, 1983; Steiner, 1979, Heines, 1995; Bryson, 2004).

12. Manager and various constituencies on organizational are involved (Lorange & vancil, 1977; Barker &Smith, 1997; Steiner, 1979, Heines, 1995; Bryso, 2004).

13. Money available to finance some of the initiatives (Lorange & Vancil, 1977; Barker &Smith, 1997; Keller, 1983; Kotler&Murphy, 1983; Steiner, 1979; Bryson, 2004).

14. Planning document (Lorange & Vancil, 1977; Keller, 1983; Meredith, 1993; Steiner, 1979; Bryson, 2004).

15. Preplanning (Steiner, 1979, Heines, 1995; Bryson, 2004).

16. Process must be clear to all involved (lorange & Vancil, 1977; Steiner, 1979; Bryson, 2004).

17. Strategy formulation (Ansoff, 1965; Lorange & Vancil, 1977; Barker &Smith, 1997; Kotler&Murphy, 1983; Meredith, 1993; Steiner, 1979, Wheelen & Hunger, 2003, Bryson, 2004).

2.5 Organizational Context

Organizational culture. The other organizational factors affect the planning process besides the dimensions of strategic planning (Wagner, 2006). The Smart, Kuh, and Tierney's work (1997) argue that culture has the important roles in improving organizational effectiveness. Cameron and Ettington (988) developed four types of institutional culture. They are: Adhocracy, Clan, Bureaucracy, Market.

Decision Making Framework. Another variable considered in this study is decision making. To get an understanding of how organization makes decisions about resources allocation, two decision frameworks are examined. One of decision frameworks is the rational-collegial and the other is the autocratic-political. Smart et al. (1997:263) provided the following definitions of two decision frameworks : 1.Rational-collegial, a decision approach in which "resource allocation decisions are the result of group discussion and consensus, based on the use of a standard set of procedures, and criteria reflecting what objectively seems best for this institution overall." 2. Autocratic-political, a decision approach in which "resource allocation decision decisions are customarily made by one individual at this institution, in a political

manner based on the relative power of those involved and without any particular pattern characterizing the criteria used."

2.6 The Linking strategic planning and corporate performance

Contingency theorists asserted that efforts to link strategic planning with performance will foster understanding of the influences of strategic planning on organizational performance under different situations, and will raise a consistent conceptualization of strategic planning characteristics and their connections to varying firm and environmental characteristics (Kukalis, 1991). Venkatraman and Ramanujam (1986) described that performance is a continuing theme in most areas of management, including strategic management, and it is of interest to both academic scholars and practicing manager. Performance enhancement is at the heart of strategic management and the importance of organizational performance in strategic management can be reasoned on three aspects, they are : theoretical, empirical, and managerial. Theoretically, the concept of organizational performance is at the heart of strategic management and the time of test of whichever strategy. Empirically, most strategy research investigations use the construct of business performance to test a sort of strategy content and process issues and for an systematic analysis of the extent to which the empirical inquiries reflect the performance dimension. The managerial importance of organizational performance is all too manifest in the many recommendations offered for performance enhancement such as research on corporate turn around and organizational transitions. Similarly, Ramanujam and venkatraman (1987) depicted that an issue central concern in the literature on strategic planning is the connection between strategic planning and organizational performance.

Numerous empirical studies have tried to explain the linkage between strategic planning and organizational performance. The seminal early works on formal strategic planning processes (e.g. Ansoff, 1965; Anthony, 1965; Steiner, 1979, Lorange, 1980, Keller, 1983) link the strategic planning to organizational performance (change). Keller, writing in 1983, presented that strategic planning places long term liveliness and quality of an organization first. It is concerned about organizational survival. Steiner (1979) depicted that "strategic planning is the systematic and more or less formalized effort of a company to establish......strategies to achieve objectives and basic company purposes" (p.15). Lorange (1980) presented that "the purpose of strategic planning is thus to accomplish a sufficient process of innovation and change in the firm.....if a formal system for strategic planning does not support innovation and change, it is a failure."

Furthermore, Lorange (1980) and Steiner (1989) explained that strategic planning system is a critical process for top management. Further Steiner pointed out that strategic planning facilitated growth. Ansoff et.al's study (1971:96) on mergers and acquisitions of US manufacturing firms 1946-1965 revealed that "the presence of planning variables examined was highly correlated with performance, specifically sales, earnings, earnings/common equity, earnings/total capital, and debt equity ratio." Further, they argued that Planning prosess mostly results in better alignment and financial outcomes than does trial and error learning. Another review of organizational performance and planning linkage (Porter, Todor, Fielding, Spendolini, & Dalton, 1980) revealed that in both the private and private sector, "organizational

performance is the single most important dependent variable." Strategic planning allows organizations to react fast in a dynamic environment, to find more alternatives, and to build new techniques and it is a key to survival, improved decision making, profits, and avoiding mistakes (Gup, 1980).

Several later works observed the criticality of performance as dependent variable, Hambrick for example argued that research question should address the linkage between strategy and organizational performance (Shrader et al, 1984). Amstrong (1982) presented that formal planning system "results in the collection and interpretation of data critical to creating and maintaining organization environment alignment." Philipps (1996) argues that in addition to strategic planning being essential to the survival of the firm, it can also contribute positively to a firm's performance. The other study focusing on the association of strategic planning to strategic change also revealed that a generative model of planning processes extends the possibility of connecting the two concepts more effectively (Liedka, 2000). Grant's study (2003) of the major oil companies in USA revealed that during the 1950s and 1960s the diffusion of strategic planning among big firms, followed interest in strategy as an area of management. Further, Grant states that Focus of empirical research in strategic planning by 1980s was on just two areas: first, the impact of strategic planning on firm performance where there were a lot of studies but no robust findings. 2. Second, the role of strategic planning in strategic decision making where it studied the organizational process of strategy formulation.

There is, however, some doubt relating to the efficacy of strategic planning to organizational performance. Scott, Michell, and Birnbaum (1981), for example, stated that ,,the research on strategic planning is inconclusive." Quin, a leading proponent in strategy.research for Logical incrementalism has concluded that the most main strategic decisions are made outside the formal planning process (1980:2). Moreover, formal planning system is often tended to be ,,a bureaucratized, rigid, and costly paper shuffling exercise." Huff and Reger (1987); Rogers, Miller, and Judge (1999) recommended that one of the reasons for the inconclusiveness of process – performance studies is the absence of strategy content. Perhaps, the most compelling critique was presented by Mintzberg (1994); Mintzberg, Ahlstrand, Lampel (2009) arguing that business had often an counterproductive love affair with strategic planning from 1960s through the 1980s. Finally, Peter (1994) made essentially the same opinions about strategic planning and other management trends such as total quality management, reengineering, and the learning organization- what Peters named "death by a thousand initiatives," (Dooris, 2002-2003).

3. Data and Methodology

This study employed qualitative research approach as framework to study (Merriam, 1998), (Willis, 2007), (Gall, Borg & Gall, 1996). Qualitative research is "an umbrella concept covering several forms of inquiry that help us understand and explain the meaning of social phenomena with as little disruption of the natural settings as possible" (Merriam, 1998:5). Qualitative research serves to explore themes and relationship at the case level (Gall et al, 1996).

3.1 Design of Qualitative Study

The research problem and sample selection.Data collection and Analysis.Writing up qualitative research (Merriam, 2002)

3.2 Research Strategy

A multiple-case study was adopted as a research strategy

Rationale:

-to investigate a contemporary phenomenon in depth (Yin, 2009) within its real life context, particularly when the boundaries between phenomenon and context are not clearly evident. Since strategic planning process is part of social structures in an organization, then, it is very difficult to separate the process of strategic planning from the context in which the process happens.

-Research on business firms has often assumed the form of case studies (Yin, 2003).

-to examine of specific phenomenon such as a program, an event, a person, a process, an institution, or a social group" (Merriam, 1998).

-to understand processes of events, projects, and programs and to context characteristics that will shed light on an issue or object" (Sanders, 1981).

-"to develop general theoretical statements about regularities in social structures and process as a reason for using a case study" (Becker, 1968:233).

-A multisite, multisource research methodology or medium grained research methodology is suggested for corporate strategy student in order to attain generalizability in reporting research findings while not losing the nuances and insights concerning individual firm's environmental context (Harrigan, 1983).

-By doing cross-case analysis a researcher will gain insight into how "processes and outcomes that occur across many cases, to understand how they are qualified by local conditions, and thus develop more sophisticated descriptions and more powerful explanations," Miles and Hurberman (1994:172).

3.3 The sample

The population of this study is the state owned banks in Indonesia (N = 5 banks) of the total of the banks in Indonesia (N =132 banks). From this population, a sample of three banks was selected (figure.2). In Indonesia some banks have formal strategic planning process and high performance, while other high performance banks have no formal strategic planning process. Conversely, some banks undertaking planning succeed, while other do not. In order to participate in this research, banks need to have:

- formal planning strategic in place

- high performance.

The first criterion was determined by:

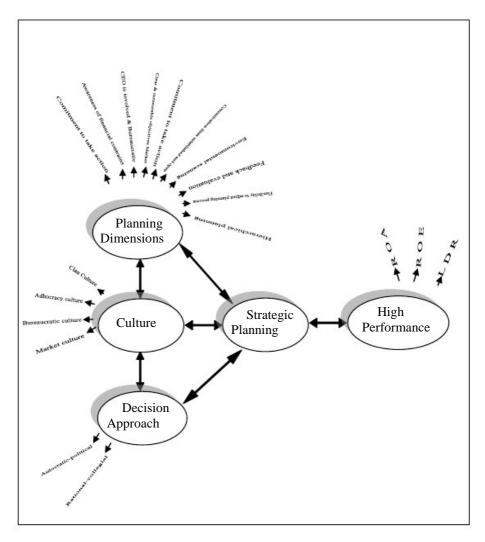
- observing the bank

- reading relevant bank's documents

The second criterion used research reports from expert panel to determine the three highest performing banks, including Journalist research reports on the performance of

the banks that annually evaluate and publish their research were used to identify the seven highest performing banks from 132 banks identified above.

Figure 1: Model of variable relations (Adapted and developed from previous Research, Wegner, 2006)



3.4 Data Gathering

Data gathering will take place at each bank of the three banks identified in the sampling process. To gather data, interviews were conducted at each bank. There are two levels interview in this study:

First, head of planning, and the other key staff. These participants are considered strategic planning team members. The Questions of interview were centered on:

- the strategic planning process

- the type of decision making approach

- the dominant culture of the bank.

Second, a group of bank staff who are not the members of planning team. The questions of interview were asked to get to know their views about :

- the strategic planning process

- the type of decision making approach

- the prevailing culture of the bank.

3.5 The data collection method employed

- In-dept semi-structured interviews of both key informants (head of planning department and other planning members) as well as non-planning members in their organization.

- Documents were analysed to validate and add to the interview data.

- The research was carried out during period June 2011 to February 2012 at three state-owned banks in Indonesia.

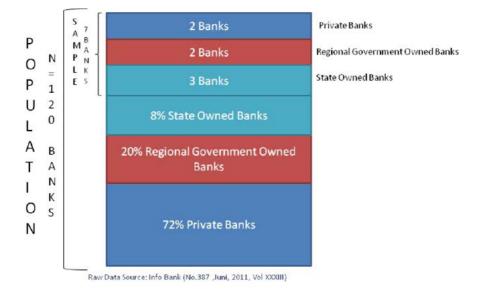


Figure 2 The population of the banks, sample, and ownership structure of the bank

3.6. Data Analysis

Analysis of case study evidence is one of the most difficult aspects of doing case studies, particularly because the techniques still have not been well defined (Yin,2009). To overcome this circumstance, "every case study analysis should follow a general analytic strategy, defining priorities for what to analyze and why" (Yin, 2009:126).

Four strategies can be used concerning with case study analysis:

- Relying on theoretical prepositions
- Developing a case description.
- Using both qualitative and quantitative data
- Examining rival explanations.

The field data also were analyzed using the theoretical framework of the model. The six generic steps as recommended by Creswell (2003) were utilized to describe the data analysis of this study (figure 3). Thematic analysis (Miles and Huberman, 1994) will be used to analysis the data.

3.7 The high-Performing state owned Banks

The first step related to this research was the determination of the banks having high performance. A total of 4 letters were sent out to the presidents of the high-performing state owned banks. Three banks agreed to participate in the study at which then interviews were scheduled. All three banks are in Jakarta. The three banks are referred to as Bank A, Bank B, Bank C. Table below summarizes the attributes of the interviews.

	Bank A	Bank B	Bank C
Interviews	6	7	7
Head of Planning Division	1	1	1
Staff of Planning Division	4	4	5
Staff of Bank	1	2	1
Males	4	5	4
Females	2	2	3
Planning Team	5	5	6
Non-Planning Team	1	2	1

Table 1 Interview Summary

4. Empirical analysis

The initial findings give researcher an understanding relating to steps, processes of planning practices, organizational culture, decision making types, organizational structures, and corporate performance at each of seven banks. The initial research results also provide insight into the contribution of organizational culture and decision making approach in building organizational context to facilitate successful strategic planning.

5. Conclusion

5.1. Research Contribution

5.1.1. Managerial implications

- a. The dimensions of strategic planning processes and cultural context are all important determinant of corporate performance. These findings have several implications for managers seeking to improve the performance in their company.
- b. For strategic planning to be effective and useful, there must be commitment and involvement all over the organization. It is very important to overcome any inherent problems such as: rivalry among divisions, departments, branches, resistance to change, resource requirement, resources allocation, and so on.
- c. The strategy initiatives and directions set up by firm management in the form of mission and vision statements and targets for cost saving, debt/equity ratios embodied as argued by Grant (2003:515) "a framework of constraints and objectives that bounded and directed strategic choices."

d. d. The connection of goals to budget, feedback and evaluation has shaped a management system much more intimately aligned with supervising efficient performance, giving incentives, and distributing cash flows to high yield usages.

5.2 Research directions

- a. To get better understanding of strategic planning practices, the future study needs to incorporate some additional variables and be examined for cross industries, private sector, non-profit organizations, and small medium enterprises.
- b. The future study needs to be examined using quantitative method with larger sample.
- c. Future research needs to consider strategy content and additional constructs such as vertical integration, operating environment, and so on.

Reference

Ansoff, H I (1965) Corporate strategy, McGraw Hill, USA

Ansoff, H I. (1990) Implanting strategic management. Prentice Hall International, UK,

Ansoff, H I, Brandenburg, Portner, Radosevich (1970) Twenty years of Acquisition behaviour in America. A comparative study of mergers and acquisitions of US manufacturing firms 1946-1965. Vanderbilt University Press, USA Anthony, RN (1965) Planning and control systems, a framework for analysis, The president and fellows of Harvard college, USA

Ardekani, MY and Haug, RS (1997) Contextual Determinants of Strategic Planning Processes. *Journal of Management Studies*, 34, 5, Pp.729-767.

Athiyaman, A and Robertson, RW (1995) Strategic planning in large tourism firms: an empirical analysis. *Tourism management*, 16, 3, 199-205.

Barker, TS and Smith,,HW (1997) strategic planning: evolution of a model, Innovative higher education, 21,4,pp.287-306.

Bryson, JM (2004) *Strategic planning for public and non profit organizations*. 3rd edition, Jossey buss, USA.

Cameron, K.S and Ettington, DR (1988) The conceptual foundations of organizational culture. In: Smart, J.C. (eds), *Higher education: Handbook of theory and research* Vol 4, Kluwer Academic Press, Dordrecht, the Netherland, pp. 356-396.

Charles, W and Jones, GR (1992) Strategic management theory. An integrated approach. Houghton Milfin Company, USA

Creswell, JW (2009) Research design, qualitative, quantitative, and mixed methods approaches, third edition, Sage, London.

Dooris, MJ; Kelley, M; Trainer, JF (2002) Strategic planning in higher education, in

Dooris, MJ; Kelley, M; Trainer, JF (*Eds*) New Directions for Higher Education no. 116, Fall, John Wiley&Sons, Inc., USA, pp.5-11.

Dooris, M.J (2002-2003) Two decades of strategic planning, *Planning for education*, issue 31,2, page 26.

Drucker, PF (1974) Management, Tasks, Responsibilities, Practices. William Heinemann Ltd, London.

Drucker, PF (1959) Long-range planning: challenge to management science, *Management Science*, Vol.5, No.3, pp.238-249.

Easterby, M; Thorpe, R; and Jackson, PR (2008) *Management research*. Third edition, Sage publication ltd, London.

Grant RM. Strategic planning in a turbulent environment: Evidence from the oil majors. *Strategic management Journal*, 24: 491-517.

Gummerson, E (2000) *Qualitative methods in management research*, Sage publication, London

Haines, SG (1995) Successful strategic planning. The systems thinking approach to building a high performance business. Axzo Press, USA

Huff, A.S and Reger, R.K (1987) A review of strategic process research, *Journal of Management*, 13, Pp.211-236.

Hunger and Wheelen (2003) *Essentials of strategic management*, 3rd edition, Prentice Hall, USA

Jarzabkowski, P and Balogun, J (2009) The practice and process of delivering integration through strategic planning, Journal of management studies, 46,8,pp.1467-6486.

Keller, G (1983) Academic strategy, The Johns Hopkins University Press, USA

Kotler, P and Murphy (1981) Strategic planning for higher education, *The journal for higher education*, Vol. 52, No.5, pp. 470-489.

Kukalis, S (1991) Determinants of strategic planning systems in large organizations: A contingency Approach. *Journal of Management studies* 28,2, pp.143-160.

Liedtka, J (2000) Strategic planning as a contributor to strategic change: A generative Model, *European Management journal*, vol 18, no 2, 195-206.

Lorange, P and Vancil, RF (1976) How to design a strategic planning system. *Harvard Business Review. September*-October, pp.75-81.

Lorange, P (1980). A conceptual approach to corporate planning, 17-73. Hartley reserve collection, Southampton.

Lorange, P and Vancil, RF (1977) *Strategic planning Systems*, Prentice Hall, Inc, Englewood Cliffs.

Lorange, P (2010), Corporate planning; an executive viewpoint, General books, Memphis, USA.

Mensah, YM and Werner, R (2003) Cost efficiency and financial flexibility in institutions of higher education. *Journal of Accounting and public policy*. 22, 293-323.

Miles, MB and Huberman, AM (1994) An Expanded sourcebook, Qualitative data analysis, Second edition, Sage Publication, London.

Miller, CC and Cardinal, LB (1994) Strategic planning and firm performance: A synthesis of more than decades of research, *Academy of management journal*, vol 37, No.6, 1649-1665.

Merriam, Sarah B and Associates (2002) *Qualitative research in Practice*, Jossey-Bass, A wiley Company 989 Market street, San Francisco.

Mintzberg, H; Ahlstrand, B; Lampel, J (2009) *Strategy Safari*, Second Edition, The Free Press, a division of Simon & Schuster, Inc, UK.

Mintsberg, H (1994) The rise and fall of strategic planning, Prentice Hall international, UK.

Norris, DM and Poulton, N (1991) A Guide for New Planners, The Society for College and University planning, USA.

Obeng, K and Ugboro, I (2008) Effective strategic planning in public transit systems. *Science direct*, 44, 420-439.

Oregan, N and Ghobadian, A (2002) Formal strategic planning, the key to effective business process management?. *Business process management journal*, vol 8 no.5,pp.416-429.

Philips, PA (1996) Strategic planning and business performance in the quoted UK hotel sector: results of an exploratory study. *Int.j. hospitally management*, vol 15, 4, 347-362.

Quinn, JB (1980) Strategies for change. Logical incrementalism. Richard D. Irwin, Inc, UK.

Ramanujam, V and Venkatraman, N (1987) Planning system characteristics and planning effectiveness. *Strategic Management Journal*, Vol. 8, 453-468.

Rogers, PR; Miller, A; and Judge, WQ (1999) Using information-processing theory to understand planning / performance relationships in the context of strategy, *Strategic Management Journal*. J. 20, 567-577.

Roney, CW (2003) Planning for strategic contingencies, *Business horizon*/ March-April. Pp.35-42.

Rosenberg, L.J and Schewe, CD (1985) Strategic planning: Fulfilling the promise, *Business Horizon*, 28,4, pp.54-62.

Rowley, DJ and Sherman, H (2002) Implementing the strategic plan, *Planning for the Strategic Plan*, 30,4, page 5-14.

Rudd, JM; Greenley, GE; Beatson, AT; and Lings (2008) Strategic planning and performance: Extending the debate. *Journal of business research*, 61, 99-108.

Schwandt, T A (1997) *Qualitative inquiry. A dictionary of terms.* Sage Publications, Inc.

Sekaran, U (2003) *Research methods for business*, fourth edition, John willey and son, USA.

Sevier, RA (2000) *Strategic planning in higher education, Theory and practice*, The council for advancement and support of education, USA.

Shrader, CB; Taylor, L; and Dalton, DR (1984) Strategic planning and organizational performance: a critical appraisal, *Journal of Management*, vol 10, no 2, 149-171.

Smart, JC (2003) Organizational effectiveness of 2-year colleges: the centrality of cultural and leadership complexity, *research in higher education*, 44, 6, pp. 673-703.

Smart, JC; Kuh, GD; Tierney, WG (1997). The roles of institutional cultures and decision approaches in promoting organizational effectiveness in two-year colleges, *the journal of higher education*, 68,3, pp256-281.

Steiner, GA (1979) *Strategic planning. What every manager must know*, The free Press, A division of Macmillan Publishing Co., Inc and Collier Mcmillan Publishers, London.

Trainer, JF (2004) Models and tools for strategic planning. New Directions For Institutional Research, no. 123, Fall, Wiley Periodicals, Inc, USA.

Tang, CY and Tikoo, S (1999) Operational flexibility and market valuation of earnings, *Strategic management journal*. 20, 749-761.

Tan, J and Peng, MW (2003) Organizational Slack and firm performance during economic transitions: two studies from emerging economy, *Strategic management journal*, 24, 1249-1263.

Trainer, J F (2004) Models and tools for strategic planning in *New directions for institutional. Research*, no. 123, Fall, pp 129-138.

Venkatraman and Ramanujam. (1986) Measurement of business performance in strategy research: A comparizon of Approaches, *Academy of Management Review*, Vol 11. No. 4. 801-814.

Wagner, RJ (2006) *Conversation on planning: Investigating the relationship between strategies, actions and performance.* Doctoral dissertation, University of Minnesota.

Willis, JW (2007). Foundations of Qualitative research, interpretative and critical approaches. Sage Publications, Inc, UK.

Yin, RK. (2003) Applications of case study research, second edition, Sage publications, Inc,UK.

Yin, RK.(2009) Case study research. Design and Methods, fourth edition, Sage Publications, Inc,UK

Role Of Leadership In Organizational Effectiveness In Albania

Dritan Shoraj¹ and Dorina Zenelaj²

¹ Department of Management, University of Tirana, Albania. dshoraj@yahoo.com

² Department of Education Science, University "Aleksander Xhuvani", Elbasan, Albania. dorina.zenelaj@gmail.com

Abstract. After 1991 Albania experienced new economic conditions and was not ready to face the demands of a capitalist economy. A large number of state enterprises and institutions were privatized while the local competition and further the global one made the organizational leaders growingly face issues requiring long term solutions. Consequently, the development seemed an impossible goal to be attained. The lack of adaptation to the scarce opportunities caused difficulties and challenges to a series of local organizations and managers. The managers, owners, employees and consumers were unprepared for this situation; therefore there was a high level of dissatisfaction.

We often speak about effectiveness but the identification of its measurement indicators for the local organizations or based only on one type of measurement indicator, seems unclear. This study will analyze the measurement indicators of organizational effectiveness under the conditions in which local managers operate. The role of leaders and managers to achieve objectives of organizational effectiveness will be attached special priority in this study.

The study will be accomplished on theoretical and empirical basis where questions will arise such as: What are the qualities and types of leadership guiding the business organizations in Albania? How has the external environment influenced and how are the opportunities for innovations and new processes managed? How successful are the female leaders in comparison to the male leaders in the Albanian business organizations? These and other related recommendations about the local leaders and managers will be further explored in this study.

Keywords: *leadership, organizational effectiveness, management, female leader, male leader.*

1 Introduction

Leadership plays a key role for the enhancement and development of a business organization. The Albanian leaders, faced with challenging prolonged transition conditions and then the global crisis, encountered numerous difficulties for the successful management of an organization. Alternatively, the traditional management methods are already deemed as outmoded and unproductive in the spirit of the overall climate of globalization and integration of business and economy as a whole. Being between the traditional and new concepts, there was even a higher level of confusion. According to Drucker, P (1998), "the main duty of the leader is to be the first violin to produce a clear sound", however it remains to be explored which would be the most appropriate sound to be produced by such a violin.

Having come from a dictatorial system where only 20 years ago, no reference was made in Albania about private enterprises and private business, the absence of a tradition and even worse, oriented to an authoritarian leadership prevailing during the dictatorship period, notwithstanding the change of the system, definitely cause difficulties to current leaders for applying new characteristics required to face transition challenges and finally the global crisis. Consequently, the analysis of this situation represents special importance for a more in-depth study of the role of the Albanian leader in business organizations. The purpose of the present study is to highlight the role of leadership in local business, its type and characteristics, as well as the analysis of hardships or challenges faced by the Albanian leadership during the period of global crisis.

Two main assumptions are raised, respectively one of them is null and void and the other one is alternative, presented as below:

 ${\rm H}_{0}{\rm :}$ Leadership plays a decisive role to enhance the organizational effectiveness of a business in Albania.

H: Organizational effectiveness is achieved through the always growing performance, increase of profit-production, cost reduction and a series of other factors which do not solely depend on the leadership performance.

Definition of key concepts: There are many definitions of "leadership" term and most of them are focused on the characteristics of a leader to guide a group of people for the purpose of achieving the goals of the organization. For instance, "Leadership is the process of making sense of what people are doing together so that people will understand and be committed" (Drath & Palus, 1994, pg, 4) or "Leadership is the ability of an individual to influence, motivate, and enable others to contribute toward the effectiveness and success of the organization." (Gary.Y, 2010, pg, 296). However, we will introduce as a definition of terms the way how we understand the main concepts of this study such as: Leadership: which intends the exercise of authority to activate the subordinates in achieving the goals of an organization or business it runs. Organizational effectiveness: achievement of goals any organization aims, leads to organizational effectiveness that may be defined as an accomplishment of the main projections for a short term future, adaptation and development for a midterm future and survival in a long-term run. Business: profit company involved in the market of goods and services.

The study hardships are linked with the fact that no studies have been available in Albania on this topic. Almost none of local researchers has assumed to conduct a genuine study on the role of leadership in the organizational effectiveness. Empirical studies about this topic are almost not implemented by local researchers; consequently, there are difficulties in drawing comparisons with other periods of time. However, on the other hand our study will be a good starting point for similar studies for other researchers, who may rely on data obtained from our empirical and theoretical study.

This study is also associated with delimitations. Ideally, these delimitations should not exist and open-ended interviews should be conducted with all local leaders. This would be practically impossible due to the long time required for the implementation of a study of such dimensions. Accordingly, the study is associated with delimitations such as: interviews will be conducted with local Albanian leaders, who mainly run the largest companies operating in the capital. The interviewees should exercise their position in private institutions and not in public ones. They should have at least 5-year job experience in organizations where they currently work and should be 30-65 years old.

2 Literature review

Studying leadership and its role in a whole organization or society is very interesting considering the fact that once the term *Leader is* used, people equal it to values such as power, dynamic characteristics of personality, intelligence and wisdom and as a good predictor of crises and collapses, either financial or social.

If we date back to the world history, we would find very special characteristics of leaders and an extraordinary influence on the life organization of a people. There suffices to mention names such as Gandhi, Mohammed, Mao Tse-tung, Julius Caesar, Alexander the Great, Adolf Hitler and the Albanian dictator, Enver Hoxha. All these leaders have demonstrated an unusual strength in leading their nations in different and often mysterious ways.

Thus, from this perspective it is clearly highlighted that a leader may bring about major changes in organizations, yet, is a good leadership the only way to achieve an effective organization? This will be also elaborated in the wake of the present study.

Leading change is one of the most important and difficult leadership responsibilities. For some theorists, it is the essence of leadership and everything else is secondary. Effective leadership is needed to revitalize an organization and facilitate adaptation of a changing environment. This subject became especially relevant in the 1980s when many private and public sector organizations were confronted with the need to change the way things were done in order to survive (Gary.Y, 2010, pg, 296). Major change in an organization is usually guided by the top management team, but any member of the organization can initiate change or contribute to its success.

On the other hand, one of the most important research questions in strategic leadership is how leaders can influence the overall effectiveness of large organizations. Even though it is clear that the influence of top executives on the performance of an organization is greater in some situations than in others, it is less clear how this influence is actually achieved. The Flexible Leadership Theory (FLT) was formulated to answer this question (Yukl & Lepsinger, 2004, pg, 34). FLT incorporates ideas from earlier leadership theories and from related areas of study, including organization theory (Katz & Kahn, 1978, pg, 650; Lawrence & Lorsch, 1969, pg, 152; Tushman & Romanelli, 1985, pg, 171) strategic management (Hambrick & Mason, 1984, pg, 193; Porter, 1996, pg, 61) human resource management (Bowen & Ostroff, 2004, pg, 206; Wright & Snell, 1998, pg, 756) and change management (Beer & Nohria, 2000, pg, 133; Kraatz & Bresser, 1999, pg, 15). The four types of constructs in FLT include organizational performance or effectiveness, performance determinants, leadership processes, and situational variables.

As Gary Yukl states in his book "Leaderships in Organizations": "Organizational effectiveness is the long-term prosperity and survival of the organization, and it depends on three types of performance determinants; 1) efficiency and process reliability, 2) human relations and resources, 3) adaptation to the environment". (Yukl, 2010, pg, 390).

It clearly seems that the leadership plays a crucial role in a business organization but it is not the only component bringing the effectiveness required for further success and stability of a business. Leaders of all levels may help to create conditions favorable to learning and innovation (Vera & Crossan, 2004, pg, 28). On the other hand, every leader should have some characteristics to achieve this objective, therefore the major change will be more acceptable and less disruptive if people develop pride and confidence in their capacity to adapt and learn. Confident people are more likely to view change as an exciting challenge rather than an unpleasant burden. To develop an appreciation for flexibility and adaption, encourage people to view all practices as temporary (Yukl, 2010, pg, 324). In addition to the foregoing, leaders should keep subordinates informed about relevant learning opportunities and make it easer for them to pursue these opportunities. Leaders can also encourage and facilitate collective learning by teams.

There is no ready-made formula on leadership. This also applies to the bestsellers such as *Leadership Challenges* of Kouzes & Posner or *The 7 habits of highly successful people* of Covey etc. The exercise of leadership includes a multidimensional integration of theory, processes and practice. What is effective in a situation may prove effective under another situation.

Understanding leadership requires multiple perspectives. During XX century the leadership was built due to intensive investigation as an area of study, mainly because the industrial revolution led to large organizations which required human resources to run them. The scientific researches concentrated their efforts to discover what type of leadership was required/needed to render these types of organizations effective. The set of scientific researches of around 90 past years, collected as a result of these efforts, modeled to a wide extent what we think of leadership. It is clear that this is not the only way of viewing things as the same leadership at the top of these organizations confronted the XXI century which caused failures for the greatest world

leaders. Then, we naturally reach the conclusion that the leader is not everything within an organization although his role is fundamental.

The economic decline highlighted some unpleasant aspects of business behavior. Business leaders, once admired, proved to be fraudulent persons and fund embezzlers; once respectable professionals were greedy dodgers; governmental organizations had poorly reacted to these phenomena. During 2002 the armed conflicts involved more than 50 nations, about ¹/₄ of the surveyed nations (National Defence Council Foundation, 2002). Furthermore, 11 September 2001 transmitted a SOS message to the developed world from the undeveloped one.

The study of leadership in the context of the range of human differences/changes is the only way of tackling this issue in the present century. The study of human changes, which is inter alia connected with races, gender, age, ethnicity, religious belief, lifestyle etc, influences almost all discussions about leadership. The leadership in XXI century is intrinsically linked with the spectrum of these differences.(Llaci, 2008, pg, 19)

In conclusion, irrespective of leadership characteristics and the role it may play in organizational effectiveness, again seems that the leadership, at least as a concept has abandoned us. Notwithstanding the great efforts of business leaders and management researchers to honor and attach it dignity, the understanding of leadership is despised due to its overuse. Leadership has been rendered powerless to keep its promises.(Sinclair, 1988, pg, 13.)

Even in the changeable world of management theory, no subject has produced more nonsense than leadership. The value of academic researches on the complexity of chaotic situations in which most of business leaders and academics find themselves, is practically zero (Micklethwaite and Woolridge, 1997, pg 37). Leadership is an intangible quality, without a clear definition. This is probably a good thing because if guided people knew the definition, they would capture and murder the leaders. Some cynical persons may admit that a leader is the one to make people do things the leader may take advantage of. Yet, it cannot be a proper definition as there are so many exclusions you are personally aware of. (Adams, 1997, pg, 122)

There is no unequivocal understanding of what distinguishes leaders from nonleaders. It has not happened to say so little and to make such strenuous efforts. (Bennis and Nanus, 1985, pg, 212) Leadership is the personal quality possessed by people which is worse defined and less understood. There are as many definitions on leadership as the number of scholars in this area. (Lipitt, 1982, pg, 69) All these definitions of the great authors on leadership clearly show the difficulty and even failure of an accurate definition of the way how a leader guides an organization, if he leads it to success on his own or in combination with other components.

3 Research methodology

This study is focused on the qualitative scientific method which is not based on the sampling of population but on data collection in specific previously defined cases. Ideally, as researchers we should manage to study all leaders in local business organizations but in a real situation it is almost impossible with this focus group. The time available to the persons selected for the study is too limited due to a series of commitments they have in business organizations where they exercise their powers.

Empirical data will have interpretative epistemological approaches, where the collected data are precisely the words and expressions of the interview participants. The investigation includes subjective connotations, feelings, thoughts, opinions and beliefs of the interviewees.

The results cannot be generalized within a broad population but their purpose is a detailed understanding of the phenomenon explored for those subjects who will be part of the study.

3.1 Research instruments

In this study, the drafting of instruments is done on the basis of the work of Holstein and Gubrium (1995). "Semi-structured interviews" will be employed, which are focused on active techniques and not with fully structured questions. The active interview is "freely oriented". It consists of both elements: objective and subjective. The objective elements focus on *what* will be voluntarily discussed between the interviewee and the interviewer. It is subjective as it influences the way how the interviewers will interpret "what the interviewers want and have the opportunity to elaborate their thought on such perception which may be even wrong". The interviews are made by keeping in line with a format of questions built by the applicants commensurate with the purpose and objectives of this research.

The semi-structured interviews include about 15 open questions allowing the interviewee to freely express himself/herself about the question posed. A *box-format* interview is used, where questions are uniformly distributed throughout the questionnaire and at the same level, except for the first demographic questions intended to indicate the employment period as a leader, age and gender.

3.2 Sampling and sampling procedures

The sampling process is associated with delimitations in the study, also identified during the first part of the study, which restricted it to the following criteria: Interviews will be made with local Albanian leaders basically running the largest companies operating in the Albanian capital. The interviewees should exercise their position in private institutions and not in public ones. They should have at least 5-year job experience in the organization where they currently work and should be 30-65 years old. The interviews with members of this focus group of interest to our study were conducted through non-probability snowball sampling procedures. The first name of a male leader and the first name of a female leader were initially selected from the list of names of the Albanian leaders invited in "The ExpoConference on Albanian Natural Resources, 3rd edition" implemented by FIAA (Foreign Investors Association of Albania. During this conference the meeting and contacts of these two first leaders selected for our study were made possible. Further, after fixing an appointment with them, receiving their consent and completing an interview, both interview participants were required to recommend us companies and other leaders sharing similar characteristics with them, in order to be further interviewed. This procedure was followed on the basis of the snowball sampling criteria, presupposing that the first interviewees would suggest persons of similar characteristics with them, with whom the contacts for interviewing procedures were developed.

Our target was to conduct a minimum of 10 interviews out of 20 we had planned for the study. We intended to reach an equal number between males and females. The sampling procedure itself aimed to ensure gender equality of local leaders. During the interviewing process only 16 interviews were made. Four of the persons recommended to be interviewed refused their participation and further interviewing. Due to the above sampling procedure 16 interviews were conducted, including 7 females and 9 males.

3.3 Ethical principles have been paid special attention in this study

Informed consent according to Terre Blanche and Durrheim (1999) and Strydom (2002) is the process of explicitly claiming the consent from the subject to be participant in the study, based on his full understanding of the interviewing procedures and its purpose. For instance, all participants of this study will be informed of the purpose of interview and the fact of keeping notes during the interviewing process. They will be allowed to read the notes at any time of the interview.

Confidentiality includes an agreement from the applicants to protect the anonymous status of participants in the study (Stydom, 2002). Very few people would like to share with others details of their private life, their opinions and emotions. Therefore, the confidentiality is a prerequisite for reliable researches (Oka and Shaw, 2000).

Result reporting: The study results should be published, attaching special importance of the rights of the participants. In no case the information provided by the study participants shall be misinterpreted or distorted (Stydom, 2002). In this study, the results are fully based on the data obtained from the participants.

3.4 Data analysis

Data analysis is made on the basis of answers received from the interviews conducted with the interviewees. Each interview is subject to the same data analysis procedures. The approach was dynamic and flexible, thus causing a return to the initial data and a potential change of analysis.

It is provided the interpretation of words, sentences expressed by each interviewee, thus drawing the meaning of data of each of the subjects. The whole process has undergone the phases of text transcription, codification and indexing. Each line of the text is labeled with ordinal numbers inserted on the right side of the text and on the other side the paragraphs were codified. Accordingly, the changes are identified, as well as the similarities and relations between different parts of textual data. These data will be tabulated, clearly demonstrating these changes, similarities or relations between defined variables. NVivo software was used to assist in data storage, organization and coding.

4 Empirical study

 Table 1: Main topics and subtopics linked with the Role of Leadership, typeand their contribution to Organizational Effectiveness

 Topics
 Subtopics

Topic 1: Personal characteristics of a successful leader

Topic 2: Leader's role in organizational effectiveness.

Topic 3: Factors enhancing organizational effectiveness.

Topic 4: Difficulties of global financial crisis and its management by leadership.

Organized, controls subordinates, discreet, patient, wise, Visionary, clearly defines his mission, meets objectives and goals. Intolerant, not emotionally involved, calm. To have a personal model and not imitate him/her, not be influenced by other leaders, to be sure of the decisions he/she takes. Sociable, rational. flexible, optimistic. High level of responsibility work, at strong character, charisma, passion for work.

Has a main role in the enhancement of organizational effectiveness. Is not the only component to further develop organizational effectiveness. The Albanian leadership is the one to determine organizational effectiveness. Is one of the factors leading to organizational effectiveness and is in no way the only factor. Only the leader may further advance a business organization.

Senior management, guidance experienced leader. Increase of production-profit, business expansion, new innovations. Good marketing, organized wellstrategy, clear objective. Increase of the number of clients, cost reduction, efficient human resources. New technology, staff qualification.

Difficulties: Limited financial resources, reduction of profit and production, unfair competition, decrease in demand, unpredictable clients, reduced human resources, psychological pressure, growing stress. Managed through utmost devotion, adaptation of the implementation of proper strategies. Efforts for keeping cost under control, flexibility with

clients.Hardly diversifying services, financial and human resources. Increase of efforts to identify new clients. By motivating the personnel, he/she overcomes any challenges, addressing issues in real time.

Topic 5: Characteristics of Albanian There are no true leaders in Albania. leaders. Thev are traditional, strict and authoritarian. overconfident. closeminded. Rather autocratic than democratic leaders. Trend of despotic leaders, leaders due to nepotism. Individualistic, little tolerant.

As mentioned in the first part of this study, 16 interviews were conducted with some of the administrators of the largest businesses in Albania who provided empirical data on the role of leadership in organizational effectiveness, his/her characteristics, as well as the global financial crisis management. Out of 16 interviewees, 7 were females and 9 were males running large businesses in Albania.

While questioned about the personal characteristics of a leader and as summed up in Table 1, most of them stated that a leader should be visionary, calm, rational, organized and patient. They generally attribute positive values to the character of a leader. Meanwhile, some 6 male leaders said that a leader should be intolerant, strict, demanding and they declared they have often used offenses and called names if their subordinates had not properly carried out their duties. On the other hand, the group of the interviewed female leaders stated that all of them have a careful and positive communication with the subordinates and are tolerant to the problems they encounter in their work but not to staff laziness or irresponsibility.

When asked to whom they attributed the failure and success within an organization, 14 interviewees said that "failure belongs to the leader and success is an attribute of the whole staff". Two of them stated that the failure and success are attributed to the leader.

Most of the interviewees admit that leadership plays a key role in organizational effectiveness but is not the only factor leading to organizational effectiveness. There are many factors influencing the effectiveness enhancement, such as increase of production-profit, cost reduction, senior management, state of art technology and new innovations, growing business, good marketing. All interviewees admit that leadership may foster organizational effectiveness of a business but is also associated with other components guaranteeing success.

The core questions about the challenging global crisis and its management brought empirical data such as: Albania is not directly affected by the global financial crisis. 2013 is expected to be a more challenging year due to the economic situation of the neighboring countries. The crisis in Albania was rather psychological than physical but it is definitely associated with difficulties to provide foreign investments, parent companies or businesses under difficult financial situation have borrowed incomes from the branches in Albania, decrease of trust and cooperation etc. Alternatively, they admitted that notwithstanding the absence of a direct crisis in the country, there have been tangible difficulties such as unfair competition, decrease of human resources, reduction of budget and profit.

This situation was managed through great efforts, devotion, perseverance, staff motivation, increase of efforts to find new clients, flexibility and well organized mid-term strategy.

While asked about the characteristics and type of the Albanian leadership, there were no changes in the answers between females and males. Almost all of them admit that the leadership in Albania is traditional. The leaders are overconfident, strict, autocratic rather than democratic. They are leaders due to nepotism and little rely on meritocracy and their individual skills. They admit that the work with foreign leaders operating in Albania is much easier and fruitful. They encounter more positive characteristics and of a democratic leader among the young leaders not older than 40 years of age but these are also few cases.

Regarding the success among female business leaders or males under the same position, according to the collected data: 5 of the females stated that females are more successful in running a business due to their tolerance, sensitivity, intuition and greater sense of responsibility. In the meantime, 11 others stated that success is bigger with a male leader because they spend more time and dedicated efforts at work. The leadership in Albania requires masculine characteristics due to the masculine society in which we live but on the other hand all males admit that it would be easier if they work with a female leader.

5 Conclusions

In the beginning of this theoretical and empirical study, our purpose was to highlight the role of leadership in the local business, its type and characteristics, as well as the analysis of difficulties encountered by the Albanian leadership during the global crisis period.

The collected data clearly demonstrate that the void assumption arguing the leadership plays a decisive role in enhancing the organizational effectiveness of a business in Albania is not applicable. In the meantime, it is proven the alternative assumption that the organizational effectiveness is achieved through the always growing performance, increase of profit-production, cost reduction and a series of other factors which do not solely depend on the leadership performance.

The interviewees clearly defined that the organizational effectiveness is achieved through a series of factors mentioned in the empirical part and not only by the role of leader within an organization. On the other hand, we reach the conclusion that the leadership in Albania has traditional features and barely democratic ones. They are authoritarian, intolerant, non-visionary. They are focused only on task division and obedience of the subordinates, lack long-term vision and are not easily adapted to the present requirements. Admittedly, there has been no direct crisis in Albania and the global crisis challenges were successfully managed. One of the most interesting data were the ones collected for the gender of the most successful leader, female or male. Most of the male interviewees were more positive to favor a female leader and the remainder, including females, believed that males are born leaders and the Albanian environment offers more opportunities for success to men than to women.

The Albanian leader should adapt his leadership style to the type of organization or business he/she runs and the situational context of the current duty. From such perspective, these factors set the limits and level of responsibility of the leadership role. Definitely, a leader plays a key role to balance clashing values and develop a consensus between the general political values of the society and the values an organization has formulated to meet its objectives. However, to achieve this, a leader in Albania should direct or guide on the basis of meritocracy, personal and professional skills, clear vision of where he/she wishes to move a business and adaptation to present requirements, place and purpose of the organization. Having the position of a leader due to acquaintances or family inherited business is not sufficient, though during recent years it seems to have changed, being considered by all businesses in Albania.

By the end of this study we would suggest the Albanian leadership pursues another approach in terms of the way how they run a business, organization or the society as a whole. Ideally, the Albanian leadership should have the characteristics of a transforming leader. Being under the influence of the globalization process, all businesses are advancing toward a global economy consisted of interlinked markets.

This evolutionary process which began only some decades ago, is currently advancing at a quick pace. Consequently, a traditional leader would not bring the development of business and further of the Albanian economy, hence, it is advisable and necessary that the leadership tradition should completely change, from an autocratic leader to a transforming leader, who shifts the focus on radical organizational change. To implement such changes, a leader should develop and "sell" a new strategic vision to the agency units and the key "stakeholders". The Albanian leadership should develop a vision based on values and organization and translate this vision into a long-lasting reality.

References

Drucker, P (1998) On the Profession of Management, Harvard Business School Publishing, Boston

Drucker, P (1999) Management Challenges for 21st Century, Harper Business, New York

Kofman, F., Senge, M. P (1995) The Heart of Learning Organizations, Learning Organizations, Oregon, Portland

Supian, W (2001) Works Sphere in USA: New tendencies 21 century, *International Journal Management problem theory and practical*. Vol, 23, 125-132

Bennis, W. and Nanus, B (1985) *Leaders The strategies for taking charge*, Harper Row, New York

Boyatzis, R. E (1982) The Competent Manager: A Model for Effective Performance, Wiley, summarised in Evan, L. and Kerrison, S.(edited by Fowler, B,1994) *MCI Personal Competence Model*, Employment Department, New York, 79

Kouzes, J.M. and Posner, B (1987) The Leadership Challenge, Josses Bass, San Francisco

Peters, J. and Smith, P.A.C (1998) Action learning and the leadership development challenge, *Journal of Workplace Learning*. Vol, 10, 145

Llaci, Sh (2008) Management, albPAPER, Tirana, Albania

Durrheim, K (1999) Applied methods for the social sciences, UCT Press, New York

Drath, W.H & Palus, C. J (1994) *Makinig common sense: Leadership as meaning-making in a community of practice*, Cennter for Creative Leadership, London

Yukl, G, (2010) Leadership in organization, Prentice Hall, New York

Yukl, G & Lepsinger, R (2004) *Flexible leadership: Creating value by balancing multiple challenges and choise*, Jossey – Bass, San Francisco

Katz, D & Kahn, R. L (1978) The social psychology of organization , Prentice Hall, New York

Lawrence, P.R & Lorsch, J.W (1969) New management job: The integrator, *Harvard Business Review*. Vol, 45, 151

Tushman, M. L & Romanelli, E (1985) Organizational evolution: A metamorphosis model of convergence and reorientation, Organizational Behavior, London

Hambrick, D. C & Mason, P.A (1984) Upper echelons: The organization as reflection of its top manager, *Academy of Manager Review*. Vol, 47, 193-206

Porter, M. E (1996) What is strategy?, Harvard Business Review. Vol,39, 61-78

Bowen, D .E & Ostroff, C (2004) The role of the "strength" of the HRM system. *Academy of Menagement Review*. Vol, 22, 206

Wright, P. M & Snell,S. A (1998) Toward a unifying framework for exploring fit and flexibility in strategic human resource management. *Academy of Menagement Review*. Vol 23/4, 756

Beer, M & Nohria, N (2000) Cracking the code of change, *Harvard Business Review*. Vol, 33,133-141

Vera, D & Crossan, M (2004) Stretegic leadership and organizational learing. *Academy of Management Review*. Vol ,29, 222-240

Gubrium, J. F. & Holstein, J. A (1997) *The New Language of Qualitative Method*, Oxford University Press, New York

Daly, J. & Engleberg, I (1994) *Presentation in everyday life: Strategies for Effective Speaking*, Houghton Mifflin, New York

Matthews, B. & Ross, L (2010) Research Methods, Pearson Education, London

Yukl, G. & Becker, W (2007) Effective empowerment in organizations, *Organization Management Journal*. Vol 3/3, 210-231

Yukl, G., Fu, P. P., & McDonald, R (2003) Cross-cultural differences in perceived effectiveness of influence tactics for initiating or resisting change. *Applied Psychology: An International review.* Vol 52, 68-82

Yukl, G., & Lepsinger, R. (2005) Why integrating the leading and managing roles is essential for organizational effectiveness, Organizational Dynamics, London

Appendix

Individual Interview: Role of leadership in organizational effectiveness in largest local businesses, May 2012, Tirana, Albania

Date of Interview:Name of Interviewer:			
Job Position			
Organization or Business he/she runs:			
Hello, my name is and I work as a lecturer at Tirana University.			
I am currently working to defend my PhD thesis. I am exploring different topics such			
as leadership and its role in the organizational effectiveness. We are two researchers,			
PhD candidates conducting investigations about this topic. We highly appreciate your			
participation in this study as you would contribute to data collection on the role of			
leadership in local businesses, challenges and determination of characteristics of the			
Albanian leadership.			
The interview will take about 60 minutes and the questions will cover topics such as your age, experience as a leader and your management experiences etc. The			
participation in this study is made on voluntary basis and you may decide if you will			
participation in this study is made on voluntary basis and you may decide if you will participate or not. If you encounter questions you do not wish to answer, then you			
may let us know. Further, you may close the interview if you do not wish to further			
continue. Though we hope you will participate in this study as your views and			
opinions are essential, again we will respect your decision.			
I should inform you that I will not share your name and your answers with anyone,			
save the research collaborator. We will not report individual answers but only the			
general answers of all leaders and managers we will talk to. The collected data will be			
used for our study which will be published abroad. Hopefully, we will talk to about 20			
leaders and managers of large businesses in Albania. I would like to inform you that			
the collaborator of this study will take note of the information you will supply during			
our interview.			
After the completion of questions and answers we would be very grateful if you			
recommended persons of similar characteristics with you, who run large private			
businesses.			
At this point, do you have any questions about our study? Do you have any questions			
about your role in this study?			
If you have nothing else to ask or add, may we continue the interview?			
YES NO			
(Mark with X the version of the interviewee, if yes further continue, if not interrupt			
the interview.)			
1) Gender? M or F			
2) How old are you?			
3) What is your present job position?			
4) How long have you been working in this position?			

5) May you describe your characteristics as a leader?

- 6) In your opinion, which are the personal characteristics making a leader successful?
- 7) What do you think about the role of leadership in enhancing the organizational effectiveness?
- 8) According to you, to whom is the success or failure of a business attributed?
- 9) What are the components enhancing the organizational effectiveness of a business?
- 10) What do you think of the phrase "think globally and act locally?
- 11) In your judgment, is the Albanian business affected by global crisis?
- 12) If yes, what difficulties have you encountered during global crisis?
- 13) How have you managed the situation during global crisis period?
- 14) How would you describe an Albanian leader? May you list some relevant characteristics?
- 15) Who are more successful, female or male leaders? May you explain?

The impact of informality at the labour market. The case of construction industry during Albanian transition period

Enida Pulaj (Brakaj)¹

¹ Management Department, University of Vlora, Email: enidapulaj@yahoo.com

Abstract. The aim of this paper is to illucidate on different aspects of the labor market and the informalities that characterize this market during the transition period. Seen from the analysis of the socio-economic factors, the transition has led to considerable social, economic, and even cultural gaps, thus creating a fertile environment for the informal economy. The analysis is based on problems regarding the labor market and the opportunity to minimize them.

Keywords: *labor market, informality, construction industry, transition, employment.*

1. Introduction

Following the political changes of 1990-1991, there were a series of economic and political changes in Albania whose sole purpose was to transition from a centralized economy to an open market economy. During this period were observed critical social problems such as poverty, unemployment, the insecurity of individual, high prices and low wages. In such a situation was the Albanian economy at the first years of the transition period. Starting from this situation we need to emphasise that the informal sector often wants to substitute and partially plays the role of the Social State. At first glance, the informal labour market and its different forms seem to be a way of creating flexibility in the usage of physical labor.

The principal aim of this paper is to make an evaluation and to analyze the issues concerning the challenges of construction industry during the transition period in Albania. The aim of this paper is to illucidate on different aspects of the labor market and the informalities that characterize this market during the transition period. Seen from the analysis of the socio-economic factors, the transition has led to considerable social, economic, and even cultural gaps, thus creating a fertile environment for the informal economy. The analysis is based on problems regarding the labor market and the opportunity to minimize them. The main goal of this study is two-fold: (1) to provide a general overview of the contributions to the literature on the informal sector, with a special focus on the transition economy; and (2) to highlight these contributions across the construction industry in Albania. The paper focuses on the criteria used to define the informal sector, the relationship between the formal and informal economy. The paper is structured in three parts:

- ✓ Informality in the labor market.
- ✓ An overview about the Albanian situation at the beginning of the transition period.
- ✓ Informality in the labor market, the undeclared work in the construction industry and the way to improve a better situation.

According the purpose of this work there are following the description and comparative methods.

1.1 Informality in the Labor Market

To address the question of 'informality' the existing literature has come up with various definitions and concepts. As De Soto (1989) emphasizes that there is no distinct border between formality and informality in the legal sense. So, there is no clear distinction between formality and informality according to a single concept. The boundary of informality and formality is extensive (De Soto 1989). The term *informality* means different things to different people such as: unprotected workers, excessive regulation, low productivity, unfair competition, evasion of the rule of law, underpayment or nonpayment of taxes, and work 'underground'(World Bank, 2007). While the literature widely recognizes the ambiguity of the term, there are two most commonly used definitions in the labor market literature. On the one hand, the productivity (ILO) definition characterizes informality in the labor market by job characteristics and on the other hand, the legalistic definition is based on the noncompliance to the state in terms of labor laws and social security systems. The *productivity definition* and *legalistic definition* of informality are the two most commonly employed in the labor market literature (World Bank 2007).

(1) The *productivity definition* characterizes informality in the labor market by job characteristics. The informal segment of the labor market under this definition constitutes non-professionals, unskilled, marginal jobs, the self-employed, domestic and family workers and workers in small firms with up to 5 employees (Hussmans 2004).

(2) The *legalistic definition* characterizes informality in the labor market by the noncompliance to the state in terms of labor laws and social security systems (Saavedra and Chong 1999). The informal segment of the labor market under this definition is characterized by workers and the self-employed non-compliant or without access to the social security system or pension system.

Beyond the discussions over definitions and terminology, there have been two central debates in the informality literature. The early literature was largely about the relationship between formal and informal sectors (Loayza, 1996). The dualist position sustained that developing countries have two distinct urban economies (Heintz &

Slonimczyk, 2007), while their critics thought the formal and informal sector were simply two aspects of the same, single, capitalist economy.

In this respect it is necessary to determine what particular form of informality is present in the labor market of a country or industry. This has wider implications for labor market studies, which analyze the labor market and the nature of informality. The debate in the literature has centered around the question whether informal and formal labor markets are segmented or integrated. In the traditional view, based on the Harris-Todaro model, informal and formal labor markets are segmented. Informal workers are 'involuntary' in the informal labor market and barriers to entry prevent them to move in the formal labor market. Contrary to this, it has been argued that 'voluntary' choice of informal workers and self-employed also characterizes the informal labor market (Maloney 1999). Here, the two labor markets are not segmented but integrated. A third view proposed by Fields (1990) combines these two views: the informal sector is seen as very heterogeneous and two-tier. Some informal sector participants are there involuntary and segmented from the formal labor market while others work voluntary in the informal sector. To test the presence of segmentation, integration or a heterogeneous two-tier informal sector suggested by Fields (1990), empirical evidence generally tries to assess (1) the existence of a wage gap and/or (2) job mobility between the formal and informal segments.

The definition of informality and the informal sector poses a challenge in itself due to its very nature of not being easily observable (Schneider and Enste 2000; Mead and Morrisson 1996). A very broad definition defines the informal economy as including "unreported income from the production of legal goods and services, either from monetary or barter transactions, hence all economic activities that would generally be taxable were they reported to the tax authorities" (Schneider & Enste 2000). The diversity in definitions of the informal sector is a result of the fact that different units of observation and different criteria of informality have been used. We can identify four main units of observation: enterprises, activities, income and people. Similarly, we can identify three main criteria used to determine informality: registration (mainly for tax and social security purposes), measurement (in GDP statistics) and regulation (mainly labour regulation). The informal sector has been defined by any combination of the above units and criteria. Thus, for instance, it has been defined as the set of all income that escapes measurement, all enterprises that escape registration, all activities that escape regulation, all income that escapes registration, all people whose work escapes registration, all enterprises that escape regulation, and so forth.

1.2 The Albanian situation at the beginning of the transition period.

Informality in the context of transitioning away from the planned economy has other interesting features. In all former socialist economies the initial period of transition was characterized by the privatization of the large state enterprises and the birth of new firms that try to exploit the profit opportunities opened up by the withdrawal of the state from the economy. These new companies are particularly important because they can provide the consumer goods and services that were rationed during the planned economy times. During the transition period in Albania were observed critical social problems such as poverty and unemployment. These initial conditions that were created made it very difficult to accept such a high unemployment and social polarization. The large industrialization scale, the absence of the market mechanisms, and the lack of the competition were typical characteristics of the Albanian market. The local economy until the last years of the transition was based on:

- ✓ The labor distribution in manufacturing and heavy industry characterized from The outdated technology that was often unproductive.
- ✓ There existed a limited experience regarding the way the free market and competition functioned.
- ✓ There was no free labor market.

The change from a centralized economy to an open economy was difficult. The early years were characterized with an immediate growth of the number of the unemployed as a result of the interrupted activity of many economic, agricultural, and industrial companies.

There was achieved the rapid privatization and market liberalization during the transition period which created an appropriate environment for the domestic and foreign investments and foreign. Labor market has changed completely during this period. One such market did not exist before. This is because the state was the only buyer of labor and the only regulator of the economy including also the labor market.

1.3 Informality in the labor market and the state of undeclared work in the construction industry

There are different opinions about the definition, the role and the size of the informal sector. Some people think that the informal sector serves as a lubricant for the market economy creating work places (Becker 2004). Other authors in Albania wrote that informal economy may be defined as "that part of the 'lawful' economy which doesn't respect the enacted legal framework and which would be rate able in case of report to fiscal authorities" (Mema and Preçi, 2003). In general terms, the informal economy refers to "all economic activities by workers and economic units that are in law or in practice – not covered or insufficiently covered by formal arrangements" (ILO, 2002). But, this sector has negative effects which serve as barriers for the economic development. The economic activity of the informal sector coexists with the structured sectors, even in many cases it competes successfully (Bello et al. 2011). Sometimes, the informal sector provides work places and the incomes for many people, increase the purchasing power, provide a short-term development and many people seek to maintain this position. But the existence of this sector has many negative sides as shown in the short and long run. The government is not the only which requires the implementation of the laws but even the unions and the employers who need a more conducive legal environment for their businesses. The regularity of the policies implementation, in the labor market, depends on the size of the informal sector. All the societies in transition pass this labor market phenomenon but in certain periods the relationship between the formal and the informal economy change.

The construction sector is the typical sector where the undeclared work is evident not only in Albania but also in Europe. In Albania, research on undeclared work in general and specifically in selected sectors are rare. But it is proved and well-known that the informal employment is proportionally increased with the distance from the capital city where the inspections are less enforcing. One of the major reason of the undeclared work in the construction and concentrated mostly in the coastal areas and in the outskirts of the major cities is construction of buildings – houses, villas and apartments blocs without construction permits. In Albania, the construction sector has a high level of undeclared work which reaches 19%, while the economic informality reaches 7-9% (INSTAT, 2009). Nowadays, this sector is one of the profitable and important sectors to the Albanian economy, about 7.5% level of growth.

There are many factors that affect the informality in the construction industry. The intensity of this phenomenon was higher especially in the early years of transition. Here we can mention some of the factors which influenced the beginning of the Albanian transition period:

- ✓ The low application of legislation regarding the financial obligations of the employers, the concealment of the real workers number, hiding the real level of income so the contributions calculation on the wage level is very low.
- ✓ The high transaction costs of doing business. The obstacles, always, exist during the business activity which they become the ground for the emergence of the informal economy. The high transactions cost of doing business in the formal sector makes the companies to direct their activities towards the informal sector where the labor cost is very low. Even though, it is spent several years of transition, the statistics rank Albania at 82 places regarding the ease of doing business, among 183 economies including small economies and some of the poorest counties (Doing Business 2011).
- ✓ The uncertainty of the property law, the core principle of developing a formal business.
- ✓ The uncontrolled demographic movement of population toward the urban areas with a higher economic level. (Banerjee, 1983)
- ✓ The income from the informal sector mitigates the degree of poverty of the individuals which receive social assistance.
- ✓ The low level of cooperation between the social partners has brought the slowly interventions in those informal activities which avoid to apply the law.

So we can say that the undeclared work and the informality phenomenon in this sector affects not only the life of the employees who work in it but also other groups of the society.

- ✓ The undeclared work makes the people completely unprotected from labor legislation, especially in cases of the accidents.
- \checkmark The people must work hard for a long time without taking the reward for the overtimes.
- ✓ These workers are always under pressure of a fear that every moment may lose their jobs.
- ✓ The delay of country's macroeconomic development because the informality reduces significantly the state budget revenues as a result of the unpaid contributions (Loayza 1997).
- ✓ The informality affects the competitive rivalry between the companies, creating a great difference between production costs and the market price.

✓ The black labor market encourages the bribes to the control institutions. The companies use different ways, how to "benefit" from the controlling bodies.

It makes very difficult the union activity, especially in the implementation of labor contracts.

1.4 What can be done about the labor market informality?

Over the recent years the Albanian Governments have taken a number of initiatives and measures to reduce the informal sector economy and informal employment. The minimization of the labor market informality in the construction industry can be achieved by the cooperation of social partners which are: the state, the employer and the employee. Above all, we have to be aware about the definitions of the main causes of this phenomenon creating a smoothing environment for the construction business in order to encourage informal activities to be formalized. Some of the ways, we can fight the informality in the labour market and also in the specific labor market of the construction industry are:

- Increasing the possibility of lending the business investment which creates new work places.
- Doing carefully steps towards the formalization because the informal company's closure bring economic and social consequences.
- The foundation and the expansion of our unions including as much as possible the employees without a contract.
- The awareness of the society about the danger and the negative effects of the black labor market.
- Encouragement of the active employment policies helping the private initiatives to cover the costs of salaries and social contributions.

The experience of a number of OECD and transition countries shows that there is no magic formula or single major solution to transform informal economic activity into formal activity (OECD, 2005). The measures against the informal economy are incorporated in the regulatory reform and managed by a state body based on the Council of Ministers Order. Furthermore, some of the measures that influence the activity of the businesses which operate overall the sectors of the Albanian economy are:

- One of the important measures in this direction was *reduction of the small business tax by half*, starting from January 2007.
- Reduction of the Social Security Contribution. According to the Law No. 9600/ July 2006 the social and health insurance contributions were decreased by 9 percent on the employer' side amounting to 32.9% from previously wage share of 41.9%. It is thought that this initiative would encourage the enterprises to declare all the employed persons and total wages and salaries assigned to them.
- Adoption of flat tax. This encouraging fiscal measure is thought to determine the future performance of the Albanian companies in terms of fair self estimation and volunteer declaration for the taxable revenues, which consequently shall result in higher tax collection.

- Channeling the wage payment through the banking system. Another measure to reduce the undeclared work in the economy was the legal obligation to execute

the payment of the through the bank channels for the employees of the private sector. This is mandatory procedure in order that these expenses are recognized by the tax authorities.

- Simplification of the business registration. Considering the dynamic growth of the small business sector, the lack of a true voluntary compliance culture and the weak tax administration capacity, simplified taxation was thought to be an important element of a strategy to address the problem of underground economy. The simplified taxation schemes has indicated at least that the tax base i.e. number of taxpayers have been broadened.

- Restriction in bid participation. In the course of restricting the level of undeclared work last year the Government took another measure to optimize the utilization of the public money through the public tenders. Apart from the procedural changes and qualification criteria improvements, the new law requires also full declaration of the number of the employees and the evidence of payment for the Social security contribution and personal income tax for the declared employees for the company that takes part in bidding process.

- Obligatory registration of the self-employed persons.
- Policy approaches against undeclared work in the construction sector.

References

Banerjee, B. (1983). "The Role of the Informal Sector in the Migration Process: A Test of Probabilistic Migration Models and Labor Market Segmentation for India", *Oxford Economic Papers*, Vol.35(3), pp. 399-422.

Becker K. F. (2004), "Fact Finding Study: The Informal Economy", Swedish International Development Agency.

Bello, M., Cucllari, F. & Cini, M. (2011). "On the development of the albanian informal economy", *Management&Marketing*, Vol IX(2), pp. 315-322

De Soto, Hernando. 1989. The Other Path – The invisible revolution in the third world, I.B. Tauris & Co Ltd Publishers, London.

Fields, G.S.(1990). "Labour market modelling and the urban informal sector: theory and evidence." Cornell University, ILR collection.

Harris, J R. and Todaro, P. M. (1970). "Migration, Unemployment, and Development: A Two-Sector Analysis." *The American Economic Review*, Vol 60(1), pp. 126-142.

Heintz, J. & Slonimczyk, F. (2007). Beyond Dualism: Multi-Segmented Labor Markets in Ghana. University of Massachusetts – Amherst. *Economics Department Working Paper Series*. Paper 35.

Hussmanns, Ralf. 2004. "Defining and measuring informal employment." International Labour Office. Switzerland.

ILO (2002), "Resolution Concerning Decent Work and the Informal Economy", International Labour Market Report.

Information data from Albanian Institute of Statistics (INSTAT).

Loayza, N. V. (1996), "The Economics of the Informal Sector: a Simple Model and Some Empirical Evidence from Latin America", *Carnegie-Rochester Conference Series on Public Policy*, 45 (1), pp. 129-162.

Maloney, F. W. (1999). "Does Informality Imply Segmentation in Urban LaborMarkets? Evidence from Sectoral Transitions in Mexico." World Bank Economic Review, 13, pp.275-302.

Mead, C.D. & Morrisson, C. (1996). "The informal sector elephant." *World Development*, Vol 24(10), pp.1611-1619.

Mema, F.and Preci, Z. (2003), "Comments on: Hidden Economy a Barrier for a sustainable Development", Fourth Conference of Bank of Albania "*The Albanian economy: Progress and Challenges of economic policy*" September 2003, Series of books on scientific conferences, pp.63-69.

OECD (2005), "The Informal Economy in Albania. Analysis and Policy Recommendations", OECD 2005.

Saavedra, J. & Chong, A. (1999). "Structural Reform, Institutions and earnings: Evidence from the Formal and Informal Sectors in Urban Peru." *Journal of Development Studies*, Vol 35, pp.95-116.

Schneider , F. &Enste, D.(2000). "Shadow Economies: Size, Causes and Consequences." *Journal of Economic Literature*, Vol 38(March), pp.77-114.

World Bank. 2007. Informality: Exit and Exclusion. Latin American and Caribbean Region. Washington DC.

The influence of gender and location on young Saudi Arabian customers' choice, satisfaction and usage of mobile phone services

¹Badr Alharbi

¹Faculty Business and Law, Victoria University

Badr.Alharbi@live.vu.edu.au

Abstract

The general population in Saudi Arabia has only been able to access mobile services and the Internet from 2004 when the sector was first opened to private industry. Since then, the use of mobile services has increased exponentially as Saudis swiftly adopted the new technology. Even in an increasingly globalising cultural environment and continuing technological advancements, the use of mobile and internet services is also affected by local cultural conservatism. This quantitative study investigates the demographic factors (location and gender) that influence young Saudi Arabian customers' choice, satisfaction and usage of mobile phone services.

Keywords: Mobile phone services, Customer satisfaction, Saudi Arabia, Gender, Location.

1 Introduction

Saudi Arabia lacks an extensive landline network, as it did not experience the evolution of telephony that occurred in other more mature economies. As a consequence, Internet service was relatively rare in homes across the nation, and where available it was slow and expensive. Information and communications technology (ICT) has long been an issue of priority on Saudi Arabia's economic Five Year Plan. A considerable amount of resources have been employed to build telecommunications infrastructure in an effort to maintain a competitive position in the global economy [1]. The government-owned Saudi Telecom Company (STC) was the sole provider of telecommunications, and the company still controls all landlines in the country. In 2004, the government took a drastic measure to further the ICT revolution by opening the sector to private industry. Privatization of mobile services brought in new providers: Mobily in 2004, Zain in 2007 and GO Telecom in 2009 [2, 3, 4]. Of these, GO Telecom has not reached the status of a national provider and is therefore not considered in this study.

Saudi Arabia ranks third in the world in terms of its mobile penetration; on average, the population has 1.88 subscriptions per head. This large number is partly due to the annual influx of 2 to 3 million pilgrims on *hajj*, a Muslim rite, as they buy temporary mobile services for use during their stay. Since this large population of pilgrims buy pre-paid services, this affects the overall national data which shows a greater preference for prepaid services over long-term contracts. The following table illustrates the growth of services over a five year period, from 2005-2010.

able 1. Glowth of Mobile Services in	Saudi Alab	ia, 2005-2010
Subscription population 28m.	2005	2010
Fixed telephone	3.8m	4.2m
Fixed internet	1.3m	>1.9m
Mobile phone	14.1m	51.5m
Mobile internet	0	>1.4m
Internet users	3m	11.4m

Table 1: Growth of Mobile Services in Saudi Arabia, 2005-2010

1.1 Significance of Research and Contribution to Knowledge

In Saudi Arabia, young people are considered the main consumers of telecommunication services. More than two-thirds of the entire population of Saudi Arabia's youth, is distributed across the three main regions of Riyadh, Jeddah and Dammam. This population differs in their usage and methods of selection of new technology. As clients, these young consumers expect companies to provide certain levels of service quality and satisfaction to maintain their loyalty. Social networking tends to be an important channel influencing the choices made by young people, where they trade information about style leaders, equipment availability, product features and service quality.

The formation of preferences among young adults will arguably set lifetime attitudes regarding the desirability of certain brands and their services. Keeping an eye on the preferences of these young consumers is an issue of prime importance for ICT companies. New mobile service providers in the national market must develop a good understanding of customer preferences to create a loyal customer base. Also, young people generally are volatile customers whose preferences change quite rapidly with trends, and this necessitates continued market research.

This research undertakes a focussed analysis of the influence of cultural factors of location and gender in the choices and perceptions of Saudi Arabian youth with regard to mobile phone services. It makes a contribution to knowledge covering several fields related to ICT—relationship marketing, customer satisfaction, service quality, brand image, emotional benefits, price and functional benefits. As the relationships between these factors are of continuing interest to researchers, the unique environment of Saudi Arabia should serve to confirm or, in some instances, disprove previous findings.

1.2 Research Questions

How do gender and location influence young Saudi Arabian customers' choice, satisfaction and usage of mobile phone services?

The analysis will answer this main research question by following two individual subquestions:

- How does gender influence young Saudi Arabian customers' choice, satisfaction and usage of mobile phone services?

- How does location influence young Saudi Arabian customers' choice, satisfaction and usage of mobile phone services?

2 Literature Review

After surveying the extant literature on market research, this study finds the field of relationship based consumer research as the most pertinent and suitable for its aims. Relationship marketing is a pioneering theory in this field developed by Berry [6] in 1983 as a means of securing customer loyalty and repeat sales. Looking over market research theories in retrospect, Berry found it incongruous that market thought and practice in the 1980s focused entirely on new sales. Berry now advocates for an integrated concept of relationship marketing based on 'the right service performed well' [6, p. 73). Relationship marketing identified many factors as indicators of customer satisfaction: service quality, image, emotional benefits, price and functional benefits. Morgan and Hunt [7] postulate that successful relationship marketing requires relationship commitment and trust. Commitment and trust may seem like volatile or subjective human emotions that are difficult to measure objectively, but Morgan and Hunt successfully developed constructs to model these variables. Relationship commitment has received widespread support from scholars such as Gustafsson, Johnson and Roos [8] in the field of ICT market research. These authors examined the impact of customer satisfaction, affective commitment and calculative commitment on customer retention, finding support for 'consistent effects of customer satisfaction, calculative commitment, and prior churn on retention.' Trasorras [9] found that returning customers tend to advocate for a firm and are less

price sensitive to a firm's offerings. In an assessment of the United Kingdom's mobile phone market, Alshurideh [3] found that purchasers of mobile voice and data packages remained with a provider if they received expected functional and emotional benefits and had a positive experience with the provider's services.

3 Proposed Methodology

A quantitative methodology is selected for this research to identify and measure the relationships between the different variables. The data for the study will be collected online from samples drawn from university students in the governorates of Jeddah and Riyadh. The choice of male and female university students for this sample was made on a two-fold basis. The first aim was to capture responses from the cohort of young Saudi citizens emerging as the most important consumers of mobile phone services, forming opinions of service providers, and discussing their choices within their groups. The second aim was to maximise the access of the research to greatest diversity of Saudi youth population for which universities proved to be an appropriate site as tertiary education is free to all Saudi citizens and young people from all over the country congregate here.

The targeted population sample for the study will be drawn from university students in Saudi Arabia, specifically Jeddah, Dammam, and Riyadh.as shown in the following table.

Tertiary Institution	Gender	Faculty
King Fahd University of	Predominantly	College of Computer Science
Petroleum and Minerals,	male	and Engineering
Dammam		
Princess Nora bint Abdul	Female	Colleges of Computer and
Rahman University, Dammam		Information Sciences, and
		Education
King Saud University, Riyadh	Male and	Colleges of Computer and
	female	Information Sciences, and
		Education
Alfaisal University	Male and	College of Science and General
	female	Studies

Table 2: Sources of student participants

King Abdulaziz University, Jeddah	Male and female	English Language Institute
King Abdullah University of Science and Technology,	Mixed	New university, first co- educational. Mathematical and
Jeddah		Computer Sciences and
		Engineering Division

The administrative authorities of the respective universities will be approached for permission for access to their students. Approximately 400 responses are required and students will be contacted by advertisements, using printed material, the internet, and mobile devices. Interested students will be asked to fill out the questionnaire for the study.

The data for the survey will be collected and analysed in a primary and secondary process. Primary quantitative data will be collected online through a questionnaire designed in English and delivered by Survey Monkey or Saudi equivalent survey software. Data collected and downloaded from the online survey instrument will be analysed through SPSS Predictive Analytics Software or equivalent that will verify validation, descriptive and inferential statistics to answer the research questions.

4 Conclusions

Any conclusion for the study is difficult to forecast at this early stage; however, the aim of this study remain lucid and clear. It seeks to investigate the impact of two cultural factors, i.e. gender and location, on customer satisfaction of mobile phone services among youth in Saudi Arabia. This research is not only significant in terms of understanding the national market for the rapidly developing mobile phone services sector in the country but contributes a focussed market research study of university students who actually form more than half of the population of the Kingdom of Saudi Arabia.

References

- Ramady, M. (2010). The Saudi Arabian Economy: Policies, Achievements and Challenges, 2nd ed. Springer, New York.
- International Telecommunications Union. ICT data and statistics. http://www.itu.int/ITU-D/ict/statistics/at_glance/KeyTelecom2010.html.
- Alshurideh, M.T. (2010). Customer Service Retention: A Behavioral Perspective of the UK Mobile Market. PhD thesis, Durham Business School, Durham University.
- 4. Saudi Com. Forecasting 37 million subscriptions by end 2010. http://www.comworldseries.com/saudi/conference/speakers2.
- Dass, R., Jain, R. An Analysis on the Factors Causing Telecom Churn: First Findings. In Proceedings of the Seventeenth Americas Conference on Information Systems (AMCIS 2011, 4-7 August), Detroit, MI.
- Berry, L. (2002). Relationship Marketing of Services: Perspectives from 1983 and 2000. Journal of Relationship Marketing, 1(1), 59-77.
- Morgan, R., Hunt, S. (1994). The Commitment-trust Theory of Relationship Marketing. Journal of Marketing, 58(3), 20-38.
- Gustafsson, A., Johnson, M., Roos, I. (2005). The Effects of Customer Satisfaction, Relationship Commitment Dimensions, and Triggers on Customer Retention. Journal of Marketing, 69(4), 210-218.
- Trasorras, R. (2008). The Relationship of Value, Satisfaction, and Loyalty on Customer Satisfaction in the Professional Service Sector. Dissertation for Doctor of Business Administration, Nova Southeastern University, Fort Lauderdale.
- 10. Aurier, P. N'Goala, G. (2009). The Differing and Mediating Roles of Trust and Relationship Commitment in Service Relationship Maintenance and Development. Journal of the Academy of Marketing Science 38(3), 303-325.
- 11.Guo, L., Xiao, J. J., Tang, C. (2009). Understanding the Psychological Process Underlying Customer Satisfaction and Retention in a Relational Service. Journal of Business Research, 62(11), 1152-1159.

- 12.Seo, D. B., Ranganathan, C., Babad, Y. (2008). Two-level Model of Customer Retention in the US Mobile Telecommunications Service Market. Telecommunications Policy, 32(3-4), 182-196.
- 13. Arnold, T., Fang, E., Palmatier, R. (2011). The Effects of Customer Acquisition and Retention Orientations on a Firm's Radical and Incremental Innovation Performance. Journal of the Academy of Marketing Science, 39(2), 234-251.
- 14.Kuo, Y.F., Wu, C.M., Deng, W.J. (2009). The Relationships Among Service Quality, Perceived Value, Customer Satisfaction, and Post-purchase Intention in Mobile Value-added Services. Computers in Human Behavior, 25(4), 887-896.
- Srinuan, P., Bohlin, E., Madden, G. (2011). The Determinants of Mobile Subscriber Retention in Sweden. Applied Economics Letters 19(5), 453-457.
- 16.Kim, Y.E., Lee, J.W. (2010). Relationship Between Corporate Image and Customer Loyalty in Mobile Communications Service Markets. Africa Journal of Business Management 4(18), 4035-4041.
- 17. Iyengar, R., Jedidi, K., Essegaier, S., Danaher, P. (2011). The Impact of Tariff Structure on Customer Retention, Usage and Profitability of Access Services. Marketing Science. 30(5) 820-836.

What is business failure? A philosophical perspective

Joni Salminen¹,

¹ Department of Marketing and International Business, Turku School of Economics, joni.salminen@tse.fi

Abstract. What constitutes a business failure? For example, it can be argued that there is no failure until the entrepreneur stops trying. Or, it can be argued that the business process is a continuous failure based on a cycle of trial and error. We can equally argue that the bankruptcy does not indicate failure; rather, ceasing the business is a rational choice of seizing a better opportunity. Or, we can argue that bankruptcy is always a proof of inefficient use of resources. Seemingly, failure can be approached from many alternative viewpoints of research and practice. In this paper, we discuss failure from a philosophical standpoint, tackling polarized perspectives of failure and their underlying logics.

Keywords: business failure, failure research, concept of failure

1 Introduction

Success and failure are often seen as two sides of the coin, so that by avoiding failure, success becomes more achievable (Rovenpor, 2004). However, survival is not a synonym for success, as individuals define success and failure in their own terms (Cope, 2010). For example, having some revenue is generally seen less of a success than having a lot of revenue – yet, for a starting company *any revenue* is a reinforcing auspice for success. Success can neither be a synonym for optimal use of resources, as a firm's performance in a competitive setting depends on the ability of other firms to leverage resources (Beaver, 2003). Therefore, there are different degrees of success and, respectively, varying degrees of failure – in a word, relativity. This paper discusses the concept of failure, its nature and boundaries. By "philosophical perspective" we refer to asking profound questions, analyzing conceptual underpinnings, and maintaining a critical eye for existing research. Our goal is to better understand the alternative views and to communicate them to stakeholder groups: researchers, entrepreneurs, and policy makers.

2 Method

This paper is a part of a larger literature review of 114 journal articles relating to business failure, presenting some outcomes from that research process. Business failure has been studied in many fields of discipline. We performed a search query in major business research databases by using a set of pre-determined keywords relating to failure. Articles for the literature review were selected based on titles and abstracts.

3 Literature review

3.1 What is failure? Event versus process

In literature, we find many polarizations relating to the concept of failure. For example, it is possible to argue that there is no failure until the entrepreneur stops trying, for example because he perceives the future gains lower than the effort of staying in business (Gilad, Kaish, & Loebl, 1985). Or, it can be argued that the business process is a continuous failure based on a cycle of trial and error (Stokes & Blackburn, 2002). Similarly, failure may be seen as a discrete event of discontinuance or bankruptcy (Peat, 2007) – or, as a process leading to any undesired business outcome with varying degree of definitiveness (Ooghe & Prijcker, 2008). By combining these dichotomies, we can even argue that *bankruptcy does not indicate failure*; rather, it is possible that ceasing the business is indeed a rational choice

determined by learning gains (Coelho, 2005), opportunity cost of remaining in the business as oppose to closing it (Blum, 1974), or seizing a more lucrative business opportunity (Theng & Boon, 1996). The specific definition of failure carries theoretical and practical importance: if we argue that bankruptcy is a sign of failure we must also concur that bankruptcy statistics are the best way of researching failure. However, it is easy to see that a) bankruptcies do not include all failed business ventures, e.g. incorporated startup projects, and b) not all discontinuances are a failure. Such is the case when there is no loss to creditors (Lussier, 1996), when the entrepreneur makes a profitable exit, or the company merges with another one. As the organizational entity ceases to exist, it is classified as a discontinued business; whereas the stakeholders consider the venture as a success. Further, spin-offs by serial entrepreneurs are desirable contingencies of discontinuing an organization in its extant form through an exit, and their number can be used to measure the learning gains of failure (Feeser & Willard, 1989).

3.2 Interpretative definition of failure

Regarding entrepreneurs, Gulst and Maritz (2009) define failure as deviation from the entrepreneur's desired expectations. This definition builds relativity inside the concept; failure becomes an interpretation as oppose to a fact. The downside of the approach is that it requires extraneous effort from an outsider, such as researcher, to find out whether failure has taken place. Additionally, as founders may interpret failure differently, similar events may be reported as a failure and non-failure depending on the case - in effect, classification attempts become difficult as no common definition for failure can be agreed upon. We can imagine a case which is judged as a failure by a highly ambitious entrepreneur - or venture capitalist, for that matter – but a success by another, perhaps a less ambitious lifestyle entrepreneur. Based on this *interpretative definition*, no decisive definition of failure is possible. Not only this, but the perception of failure or success can change over time as the entrepreneur's mind develops, not remaining stable even within the individual (Beaver, 2003). Clearly, whether failure took place or not becomes a question of goals and motives of the entrepreneur and, if taken into account, the views of different stakeholder groups who may simultaneously judge the venture both a success and a failure. For example, investors typically see failure in terms of negative return on investment (Zacharakis et al., 1999), whereas entrepreneurs may apply different measurements, such as achievement of personal satisfaction. Even within a same group of people, there may be opposing views. This pluralism of interpretation causes headache for the researcher: Which interpretation is the truth? Finally, the researcher has his own theoretical bias: for example, organizational theorists are interested in finding different explanations to failure than scholars of strategic management, accounting, or marketing (Rovenpor, 2004). Research of failure is cross-disciplinary, united not necessarily by the same theoretical underpinnings, but a shared interest in the phenomenon.

3.3 Failure of ventures and failure of people

Clearly, we must separate the failure of a venture from the failure of the adventurer because their lifetimes are not equal. One can argue there are no failed entrepreneurs, only failed ventures (Gulst & Maritz, 2009). In contrast, the entrepreneur can be seen failing when he stops trying without success. When the probability of future entrepreneurial activities is low after failure, it is justified to speak of failure of the entrepreneur. Nevertheless, the separation of the individual from the event is a useful one. The personal benefits of failure relate to learning: When a failure, e.g. bankruptcy - if applying the bankruptcy definition - has taken place, we need to assess the quality of learning to determine the degree of failure. From an economic perspective, the costs of failure need to be weighed against the learning gains which will ideally materialize as a higher chance of success in future ventures of the entrepreneur. If the failure case leads to the entrepreneur starting a company of great success, the *return on failure* has in fact been extremely positive. If no such gains materialize, the value of learning is low. In fact, it is crucial for the society that entrepreneurs recognize more lucrative opportunities and are able to act upon them, even if this results in discontinuance of slow-moving projects in the favor of higher risk, higher gains. As there is no known formula for accurately recognizing the value of opportunity, any such learning and experience that helps entrepreneurs in developing practical capability of estimation is a desirable feature of failure. Shifting resources to more profitable allocations is a requisite for dynamic economy, even if it is seen as a failure a micro-economic perspective (Fredland & Morris, 1976). However, it is obvious that there are sunk costs of failure associated with a venture that did not perform as expected, which need to be weighed against the benefits for society and the individual.

3.4 Problems of explaining failure

In searching for the reasons for failure, one faces issues. First, the attribution problem, or, how to assign a cause for failure? There are several obstacles to attribution. For example, the reports by founders or managers may be biased – either intentionally or unintentionally presenting a distorted view or leaving out information. The typical reporting biases, such as the recall bias (Podsakoff & Organ, 1986), are applicable to survey research, often used to find correlations between various internal and external factors and failure. Founders may be unable to understand or articulate causality (Bruno et al., 1987), and sense-making involved in the process is vulnerable to biased interpretations due to complexities of the business process, such as long timeframe, several participants with potentially opposing views, and unknown factors residing outside the company (e.g., customers' interpretations). The credibility of explanations is further affected by the interpretation problem: the researcher interpreting the interpretations of the informant (Smith & Osborn, 2008). Third, there is the classification problem: How do we classify reasons for failure? When going deep into the rationale of failure, we can, for example, narrow all causes of failure down to financial reasons: The venture did not produce sufficient returns, and was cancelled as a consequence. Or, if we apply a perspective of human action: All

reasons of failure relate to the shortcoming of team/managers, because it was their choices that lead to the failure. Should they have chosen otherwise, the firm could have continued without failing. Hence, failure becomes a consequence of entrepreneurial actions. Further, in most cases we can reduce the root cause to demand: There was not sufficient market demand, so the company had no chance to survive, regardless of strategic choices. But how can we separate the market from the product? Clearly, if there would have been a different or better product, there would have been demand for it – after all, there is always demand for *something* in the market. Thus, it is equally right to say all failures result from the shortcomings of a product, just as it is to argue failure was due to lack of market demand.

Finally, communicating the classification to other researchers is troublesome: general categories such as "Management incompetence", "Bad product", or "Wrong timing", can be highly contestable, hard to verify, and inadequately defined by the researcher (Beaver, 2003). It is imperative that the researcher clarifies what he means e.g. by "Management incompetence", why it was chosen in the classification scheme, and what are its boundaries – what is *included* and *excluded*. Further, what is the role of such classifications in attribution of failure? Can we, for example, argue that management caused 50% of failure, product 30%, timing 15% and the rest was a result of unknown factors? If not, are we trapped in half-way from the goal of discovering the *veritable cause* of failure? These questions are critical when profoundly thinking the research motive, but remain unanswered in many works.

3.5 Failure and causation

The aforementioned multidimensional complexity of failure (Bruno et al., 1987) results in multiple causation, or competing explanations, thereby raising the question: Can we create mutually exclusive and exhaustive categorizations of reasons for failure? And if not, why are so many researchers attempting to do so? Further, there is a risk that by attempting to create universal classifications for reasons of failure we lose many interesting, contextual explanations. These explanations could help us understand the business case in its particular environment of time and place; not losing the environmental complexity but trying to derive the particular logic of failure from it. As such, although the question "Why did the firm fail?" remains, the sought answer is very much different: Instead of saying "Because there was no demand" we would go deeper in finding the root cause as a function of multiple factors such as the business model and customer insight, and find the failure was "Because the company X charged a price Y which was found too expensive by the customer group Z". At the same time, we must acknowledge the limits of our ability to explain failure exhaustively. The most we can do is to offer a partial explanation, which is often influenced by interpretation of the informants as well as our researcher's biases. Obviously, researchers can drill deeper into the complexity of reasons to make a contribution in the failure literature. For example, in the case of financial reasons we can discuss following aspects: 1) timing of capital, 2) nature of capital, and 3) source of capital. Timing refers to e.g. undercapitalization in the beginning of the business

process, the nature of capital means that the failure may incur due to negative cash flow (cash flow failure), lack of assets, or simply due to inability to accumulate revenue at any period. Third, the source of capital can introduce harmful limits and restrictions - for example, the deviating motives of venture capitalists and business founders are well documented as a source of tension and failure. Further, we can increase complexity to market-related factors by separating a static, permanent lack of demand – where customers simply didn't want the product – and a dynamic, temporary lack of demand; for example, an art dealer failed because during a recession consumers cut their spending on luxury goods. As such, we have to adjust for time, not only for the outcome of failure, but for other variables as well. This means that research designs capturing a snapshot of the world are by default less effective than longitudinal studies incorporating multiple perspectives, such as contrasting views of customers and the firm, both of which constitute of number of individuals. Finally, team-related failure factors can be reduced (or attempted to be reduced) to personality traits, skills and competences of the people working for the firm (Beaver, 2003). In conclusion, we argue that failure factors are arbitrary in many senses, and their utility depends greatly on the researcher's ability to treat them as meaningful constructs.

3.6 Failure and truth

We have already touched the question: What is the truth behind failure? Positivists would argue interpretations of stakeholders can be distinguished from the objective truth revealed by measuring *relevant variables* separately from the reported causes prone to sense-making, and post-hoc rationalization (Zacharakis et al., 1999). The question of truth is a part of a larger debate of *perception of facts* and *measurement of* facts. For a researcher, it may be difficult to explain failure by objective measurements, as he is forced to tackle a complex combination of business process, organization, and the external environment. To make convincing arguments, he needs to reconcile these views, which complicates research designs. It is no wonder, then, that the attempts seen in the literature are often less ambitious in their goal of discovering the objective truth behind failure. Even when such an attempt is made, the results risk being influenced by implicit assumptions. So, the question arises: What are the relevant measures to include in the research design? As researchers, we are facing a large number of hidden variables outside our knowledge and conception, which hinders our ability to explain the difference of business processes leading to failure and those resulting in success. Therefore, should we aim at creating a general theory of failure at all, or instead study the entrepreneurial processes leading to discovery of winning combinations through an iterative, continuous cycle of failure in dynamic markets?

3.7 Determinism versus freedom of choice

As noted by Augier and Teece (2008), we may think failure as a result of path dependency, or determinism, which is "so strong that the enterprise simply cannot

adapt". According to this view, the choices affecting success or failure are made at the business is started, and there is little to do afterwards. The contrasting view states that strategic decisions and dynamic capabilities, or actions of free will, shape the future of the firm (Augier & Teece, 2008). This philosophical choice affects many important questions. Can we influence failure at all, or is it granted by exogenous factors? Do initial choices determine our fate, or can we change the course of future by our own action? If we assume freedom of choice, when are the decisive choices made in the business process? In other words, do initial parameters affect more than later adjustments? The obvious answer seems to be that both views have some truth to them. It is fair to assume the existence of some path dependency, perhaps of stochastic nature, implying that, by average, certain settings and strategic choices can lead to worse outcomes than others. However, managers are given the opportunity to make corrective adjustments (Augier & Teece, 2008). This ability may be bound to internal variables to a major degree, as it can be seen difficult for a company to shape its external environment to a significant degree.

For research purposes, it is relevant to discover what combinations of internal and external factors produce better results than other, and which have a higher chance of resulting in failure. For example, Larson and Clute (1979) argue for the existence of failure syndrome which refers to a combination of personal characteristics, managerial deficiencies, and financial shortcomings possessed by certain individuals who are doomed to fail. In another vein, new ventures are often seen to exhibit high mortality (e.g. Rovenpor, 2004). The explanation to specific empirical findings may result in interesting theoretical discussion and explanatory constructs, such as liability of newness (Stinchcombe, 1965), arguing that new companies fail more often because they have less experience to solve emerging challenges and fewer resources to endure economic fluctuations and financial setbacks, as well as hire better managers and acquire customers (Rovenpor, 2004). Further, start-up companies have a tough time introducing novelty and change in the market, especially when the customers are used to existing solutions and refuse to change their behavior or break relationships with existing providers. Since new ventures, such as start-ups, tend to have little capital, they may find it difficult to compete for awareness against the existing players equipped with large war chests, market experience, and an array of defensive marketing tactics, such as predatory pricing. As we can see, a fairly tight focus offer grounds for interesting explanations; however, this does not suggest dogmatism. Alternative explanations should be presented - for example, in the previous case one can argue that start-ups are less likely to fail than mature organizations, because their founders are highly motivated and committed, they offer a "better way of doing things", and the company has initial resources, such as venture capital, to endure competitive pressure (Fichman & Levinthal, 1991). The risk of failure increases only after losing the initial advantages (Rovenpor, 2004), while later stabilizing through organizational routines and market position (Bruderl & Schussler, 1990).

4 Conclusions

4.1 Implications for researchers

Failure researchers have a shared interest in failure instead of use of same methods and focal points; they are using different lenses to study the topic. In fact, a unified theory of failure is missing from the literature. Regardless of the discussed relativity, the end goal of any business is success, not failure - even if only in one's own terms. There are limits to relativity: An entrepreneur who constantly fails is not considered the ideal type, even if he "explains away" the failure - in this case, he is selfdelusional, although the difference to rational entrepreneur is hard to make since it requires analyzing the extent and quality of learning from the process of failure. Failing involves *failure costs*, such as loss of resources (e.g. time), or emotional and financial $costs^1$. The academic debate therefore has a normative nature: How to help the entrepreneurs and managers to prevent failure and achieve success? Or, from economic perspective: How should businesses optimally allocate their resources for either profit maximization or maximization of social welfare? The tension between success and failure leads to further implications, such as: Should the entrepreneur/manager try to avoid failure at any cost, or treat it as a learning experience? How should policy makers approach failure? And: What is the role of the researcher – giving advice in an attempt to change reality, or keeping distance and examining failure through a magnifying glass?

Research on failure, such as any other research of human activity, is filled with counter-examples and opposing views, essentially deriving from multiple realities and their interpretation. As previously argued, benefits of failure research vary from micro-level (strategic choices) to macro-level (policy choices). The researcher should choose at which level to contribute, since the needs of the two groups are very different from each other. For example, the *externalities of failure* vary between the levels of examination. Based on the Schumpeterian view, entrepreneurs take risks and aim at disruption. Driven by disruption, the failure rate of new ventures is inherently high. But the goal of disruption is nevertheless beneficial for the society, because the rare successes make the creative destruction that leads to new industries, employment and innovation gains. As such, a high number of new ventures (followed by a high number of failures) is a sign of a healthy, dynamic economy (Graham & Li, 2002). In contrast, an economy with little risk taking has less new ventures and only "safe bets" emerge, leading to a lower degree of innovation. This said, the policy makers tend to be interested in lowering the failure rate of new ventures (Storey, 1994), even if there is no indication of the long term benefits of such activity. The contrasting view, then, is the individual tragedy and trauma of failure (Singh et al., 2007). What is good for the society (high activity of new ventures) is less good for founders of new ventures since they need to work in an economy of high failure rate - for policy makers, the

¹ Although, for example, a socially oriented firm emphasizes financial measurements of success less than the classical, profit-oriented firm.

solution may be to reduce the gravity of consequences at an individual level, or otherwise co-aligning the objectives of the society and the entrepreneur.

Due to differences in sampling and interpretation, the identified failure factors are not always compatible with one another. Rather, different authors give different names to the same phenomenon, and there is no generally accepted classification of failure factors (Lussier, 1996). However, even with overlapping explanations of failure there lie many problems. Although findings relating to composition of failure may lead to accurate aggregate results, or a snapshot of *situation at large*, without understanding the mechanisms and logics at work, the complexities of specific exogenous (e.g. sectors and industries) and endogenous (e.g. organizational) factors can easily distort the generalizability of results. Failure researchers also need to position their research in regards to success. Studies comparing failure and success within and across comparative samples are still scarce. However, further complexities arise, such as: What criteria to use in assessing the comparability of cases? By understanding environmental logics, such as industry realities, researchers can provide explanations that are accurate despite of necessary relaxations of absolute scientific reliability, such as not requiring the use of statistical data. In other words, the explanation of results is more important than absolute accuracy of the method.

Where should the focus of failure research be? If failure is *inevitable* and *process*, should we not consider *curing* failure rather than finding ways of preventing it? This would mean taking corrective measures inside the business process before the ultimate failure takes place, whereas prevention would mean creating such premises that decrease the chance of failure, such as better institutions, financing or education. In other words, we can spot a tradeoff between failure prevention ex ante, and dynamic process of curing failure as it is happening. This relates to our decision of seeing failure either as an event or as a process. For example, controllers would follow financial metrics; marketers would track customer satisfaction and market changes, and so on. In summary, researchers need to consider the choice of failure concept, which represents the researcher's philosophical approach to failure. Then, they decide the level of contribution and verify the consistency between method and the desired contribution. Researchers have several options. They may treat failure as a strict, easily definable concept (such as bankruptcy), or they may apply a pluralistic definition and face conceptual complexity. They may relax on the decisiveness of the method in exchange for theoretical insight, or they may strive for methodological perfection and accuracy, perhaps at the cost of interpreting the results. Finally, researchers may consider the trade-off between academically sound and pragmatically useful research. When the researcher's goal is to produce knowledge to help managers and entrepreneurs to avoid or prevent failure, the research should be designed to support this type of contribution. Large data sets may provide useful insights for policy makers, but strategic decision-makers may struggle with oversimplification in the results. As noted by Beaver (2003), some works lack the practical utility, or even theoretical contribution, despite being methodologically sound. For failure research, which is essentially driven by normative focus, this represents a serious concern.

4.2 Implications for entrepreneurs and managers

The risk of failure is substantial especially when certain conditions, such as novelty in product or market, little experience and lack of capital inside the firm, and unproven business model are met. Despite of many potential hazards, founders should boldly venture out. As there is no formula for success, the gained experience from failure provides excellent tacit information of the specific market context, otherwise very difficult to obtain. In answering the introductory question on how decision-makers should treat failure, we can distinguish risk-aversive tendencies (say, managerial type) from risk seeking behavior (entrepreneurial type). Seeking to discover what works and what not seems like a desirable trait of the ideal business man, coupled with the ability to face unpleasant truths instead of detaching oneself from business realities; a condition of entrepreneurial self-deception. At a theoretical level, we can define the existence of a failure-seeking decision-maker who, paradoxically, seeks to fail, although not in a self-destructive way, but for learning gains, excitement and other motivational incentives (Cope, 2010). He recognizes failure as not a permanent state of affairs, but a passing moment in time that can be changed at any point by creating a successful venture, thereby recovering momentary "wasted" time and efforts through deferred success. Finally, he must be a very rare case: because failing results in emotional distress and is an unpleasant experience, it requires a special personality to not only tolerate failure – as oppose to behavioral avoidance of expected failure - but to actively seek it in hopes of learning gains and postponed, future benefits. Individuals who are risk tolerant and "with a plan" may correspond to this theoretical type in reality, but one can also try and change his mental approach to failure to match the idea of the *failure seeker*.

The commonly stated rationale of attributing reason to failure is that by knowing the reasons it is possible to take preventive action (Abdelsamad & Kindling, 1978). However, the relationship between knowing the cause and preventing it from taking place is not without a problem – there is a need for *strategic transition* that goes deeper into understanding the particularity of the business process at hand. For example, knowing that a lack of capital causes business mortality has little value per se, but connecting that insight to strategic solutions is a step forward in the chain of prevention. However, when most research efforts are put to finding out the cause instead of finding creative solutions, the strategic dimension can easily become neglected. From a managerial perspective, generic explanations risk to lose their explanatory power under particular contexts - in other words, knowing that management shortcomings are a reason for failure does not necessarily be useful for the manager when he needs to make strategic choices in his own business environment (as oppose to policy makers who need different kind of information, perhaps of aggregated type). Individual firms are vulnerable to different failure factors since they operate under unique conditions with different set of resources. Managers and entrepreneurs should bear this in mind when interpreting results of research, and consider if they can find sufficient structural similarity between particular research and their own case (Lukka & Kasanen, 1995).

Failure of business ventures may be inevitable, or at least highly likely. From a macro-economic viewpoint, failures are bound to take place. There is still room for interpretation: Failures can be seen as casualties in the road for disruptive innovation that is characterized by great uncertainty, but which results in benefits for the society (e.g. increase in productivity, employment and competitiveness). Or, failure can be seen as mal, something to avoid at any cost. The negative culture of failure may reduce risk taking and cause stigma for failed entrepreneurs. For the society at large, it is important to recognize the significance of failure, which derives from competition, renewal of industries and the creative destruction (Schumpeter, 1942). Avoiding failure at any cost should never be in the agenda of policy makers. For example, the short-term benefits of bailing out organizations that are "too big to fail" may seem lucrative due to saved employment and reducing immediacy of e.g. financial crises, but this leads to a false sense of security. First, the society faces to opportunity cost of lost innovation: Because dominant organizations are effectively protected from bankruptcy by the state, start-ups have less chance of creating disruption that would challenge them. The incentive is given, then, to defensive strategies, not strategies of creative disruption. In the long-term such a society may face innovation loss - losing competitiveness in comparison to economies with a lesser degree of corporate protectionism.

References

Abdelsamad, M. H., & Kindling, A. T. (1978). Why Small Businesses Fail. SAM Advanced Management Journal, Vol. 43, No. 2, 24–32.

Augier, M. & Teece, D. (2008). Strategy as evolution with design. *Organization Studies*, Vol. 29, No. 8–9, 1187–1208.

Beaver (2003) Small business: success and failure. *Strategic Change*, Vol. 12, No. 3, 115–123.

Blum, M. (1974). Failing company – Discriminant analysis. *Journal of Accounting Research*, 12(1).

Bruderl, J, & Schussler, R. (1990). Organizational mortality: The liability of newness and adolescence. *Administrative Science Quarterly*, Vol. 35, 530–547.

Bruno, A. V., Leidecker, J. K., Harder, J. W. (1987). Why firms fail. *Business Horizons*, Vol. 30, No. 2, 50–58.

Cardon, M. S., Stevens, C. E., & Potter, D. R. (2011). Misfortunes or mistakes? *Journal of Business Venturing*, Vol. 26, No. 1, 79–92.

Coelho, P. R. (2005). Learning from Failure. American Journal of Business, 20(1), 1.

Cope, J. (2011). Entrepreneurial learning from failure: An interpretative phenomenological analysis. *Journal of Business Venturing*, Vol. 26, 604–62.

Feeser, H. R., & Willard, G. E. (1989). Incubators and performance: A comparison of high- and low-growth high-tech firms. *Journal of Business Venturing*, 4(6), 429–442.

Fichman, M., & Levinthal, D. A. (1991). Honeymoons and the liability of adolescence: A new perspective on duration dependence in social and organizational relationships. *Academy of Management Review*, Vol. 16, 442–468.

Fredland, J. E., & Morris, C. E. (1976). A cross-sectional analysis of small business failure. *American Journal of Small Business*, Vol. 1, No. 1, 7–18.

Gilad, B., Kaish, S., & Loebl, P. D. (1985). A theory of surprise and business failure. *Journal of Behavioral Economics*, 14(1), 35–55.

Gulst, N., & Maritz P. A. (2009). Venture failure: Commonalities and causes. *AGSE Conference*, February 2009, Adelaide.

Lukka, K., & Kasanen, E. (1995). The problem of generalizability: anecdotes and evidence in accounting research. *Accounting, Auditing & Accountability Journal*, Vol. 8, No. 5, 71–90.

Lussier, R. N. (1996). Reasons why small businesses fail: and how to avoid failure. *The Entrepreneur Executive*, Vol. 1, No. 2, 10–17.

Lynford , E. G., & Chunyan, L. (1997). Cultural and economic influences on current accounting standards in the People's Republic of China. *The International Journal of Accounting*, Vol. 32, No. 3, 247–278.

McGrath, R. G. (1999). Falling Forward: Real Options Reasoning and Entrepreneurial Failure. *The Academy of Management Review*, Vol. 24, No. 1, 13–30.

Ooghe, H., & Prijcker, S. D. (2008). Failure processes and causes of company bankruptcy: a typology. *Management Decision*, 46(2), 223–242.

Peat, M. (2007). Factors Affecting the Probability of Bankruptcy: A Managerial Decision Based Approach. *Abacus*, 43(3), 303–324.

Podsakoff, P. M., & Organ, D. W. (1986). Self-report problems in organizational research: problems and prospects. *Journal of Management*, Vol. 12, No. 4, 531–544.

Rovenpor, J. (2004). Explaining the E-Commerce Shakeout: Why Did So Many Internet-Based Businesses Fail? *e-Service Journal*, Vol. 3, No. 1, 53–76.

Schumpeter, J. A. (1975) [orig. pub. 1942] Capitalism, Socialism and Democracy. New York: *Harper*.

Singh, S., Corner, P. and Pavlovich, K. (2007). Coping with entrepreneurial failure. *Journal of Management & Organization*, Vol. 13, 331–344.

Smith, J. A., & Osborn, M. (2008). Interpretative phenomenological analysis, In: Smith, J.A. (Ed.), *Qualitative Psychology: A practical guide to research methods*, 2nd edition. Sage Publications, London.

Stokes, D., & Blackburn, R. (2002). Learning the hard way: the lessons of ownermanagers who have closed their businesses. *Journal of Small Business and Enterprise Development*, 9(1), 17–27.

Storey, D. J. (1994). Understanding the Small Business Sector. New York: Routledge.

Theng, L. G., & Boon, J. L. W. (1996). An exploratory study of factors affecting the failure of local small and medium enterprises. *Asia Pacific Journal of Management*, 13(2), 47–61.

Zacharakis, A. L., Meyer, G. D., & DeCastro, J. O. (1999). Differing perceptions of new venture failure: A matched exploratory study of venture capitalists and entrepreneurs. *Journal of Small Business Management*, Vol. 37, No. 3, 1–14.

Information and Communication Technologies

An Engineering Perspective on Emergence

Konstantinos Rousis¹, George Eleftherakis², and Anthony J. Cowling³

 ¹ South-East European Research Centre, 24 Proxenou Koromila, 54622 Thessaloniki, Greece konrousis@seerc.org
 ² Computer Science Department, City College, International Faculty of the University of Sheffield 3 Leontos Sofou, 54626 Thessaloniki, Greece eleftherakis@city.academic.gr
 ³ Computer Science Department, University of Sheffield Regent Court, 211 Portobello, Sheffield, UK A.Cowling@dcs.shef.ac.uk

Abstract. Emergence is a central element in discussions of complex systems, complexity science, and multi-agent systems. While there is a renewed interest in emergence, researchers have not yet agreed to a common view or definition. The debates seem to be endless regarding observers, the elements of surprise and novelty, and different levels, scales, and hierarchies, to name a few. This work provides a critical perspective on various emergence definitions as well as the intrinsic topics surrounding emergence, from an engineering perspective. Although the term engineering emergence might be oxymoronic according to some definitions, there are others that could sufficiently support such a goal. The critical discussion is followed by suggesting a literature definition which could serve as the basis of future work on engineering emergence.

Keywords: emergence, engineering emergence, complex systems, multiagent systems

1 Introduction

The advent of the Internet introduced an exponential increase in technology's penetration to everyday life. While less than three decades ago digital networks could only be found in the army, large universities and research organisations, nowadays the vast majority of the households in developed countries are connected on the Internet. Recent advances in mobile devices, such as laptops, netbooks, tablets, and smartphones, only increased the number of interconnected nodes per household or person. Moving further, the first house appliances, including refrigerators and microwave ovens, which are able to connect on the Internet and perform online services on behalf of their users have already launched in the market.

From a networks perspective, the world can be viewed as a huge overlay network consisting of billions of nodes interacting through a shared channel: the Internet. Yesterday's emerging trends are becoming today's norms, for example the mobile Internet or the "Internet of things". This results in a constant increase of complexity: devices have to be managed, maintained, adapted, re-configured, and so on. Classical engineering has successfully tackled these issues for many decades but as the complexity increases so does the effort required to manage this complexity. There is a growing concern that in the very near future complexity will grow beyond the point of being manageable with the traditional engineering processes that have been developed so far.

There is a common and central element in discussions of complexity science and complex systems: *emergence*. A lot of effort has been put by academia on exploring the concept of emergence and how simple interactions at a lower level of organisation can yield complex phenomena at a higher level. Observation of natural systems has demonstrated that the process of emergence is accountable for various biological properties such as survivability and adaptability. By turning to nature for inspiration, researchers have been able to reproduce similar emergent properties in artificially engineered systems including self-adaptability, self-healing, and other *self-** properties.

There is a growing concern that a point will be soon reached where an explicit design and implementation of complex systems will become impossible. At the same time, it is apparent that harnessing the power of emergence and incorporating it in a disciplined and intentional manner into systems' design and engineering, could greatly assist with the management of complexity. Directing, controlling (even partially), or in the extreme case *engineering emergence*, however, could be extremely hard and even oxymoronic according to most definitions of emergence.

Consequently, it is essential to form a thorough understanding of emergence and its intrinsic issues before exploring any such possibility. Being a topic of great debate among researches, there is a multitude of proposed definitions for emergence, spanning different disciplines, but a universally accepted definition is still lacking. This paper examines various of these definitions from an engineering perspective and concludes to a working definition, proposed by Wolf and Holvoet [1], as suitable to support future work on engineering emergence.

Section 2 provides a historic overview of emergence, with scientific attempts to define emergence presented in section 3. In section 4, emergence and the recurring topics surrounding it are discussed from an engineering perspective and a working definition is proposed for future work on harnessing emergent phenomena. Finally, section 5 concludes this paper with a critical discussion.

2 Understanding Emergence

In the complex systems and multi-agent systems fields, emergence typically refers to a *global* (or macroscopic) level behaviour which is either impossible or highly unlikely to be predicted by observing the behaviour of the individual entities at the *local* (or microscopic) level. A typical example is the flocking behaviour of birds, whereby birds fly in a lockstep formation and by analysing the behaviour of each individual bird it would be rather unlikely to predict such a globallevel result. Another example is the ability of ants to establish the shortest possible route between their nest and the various sources of food, even when the environment changes, while the behaviour of each ant (i.e. the local or micro level) is rather simplistic and by simply observing it, it would be hard to predict such a global behaviour a priori.

Lexically, *emerge* is derived from the Latin word *emergere* and according to the Oxford English Dictionary is defined as something moving out of something else and becoming visible or apparent. The same source defines *emergence* as the action or process of emerging while *emergent*, in a philosophic context, is defined as a property which arises due to complex causes and can not be analysed solely on its own or by its effects.

This last, philosophic, definition of emergence can be traced back to Aristotle who at the 4^{th} century BC wrote [2]:

In the case of all things which have several parts and in which the totality is not, as it were, a mere heap, but the whole is something beside the parts, there is a cause [...]

This excerpt is commonly interpreted in the literature as "the whole is greater than the sum of its parts." More than two millennia after, the father of modern economics, Adam Smith, wrote in his seminal work, *The Wealth of Nations* [3]:

He [the individual] generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it. By preferring the support of domestic to that of foreign industry, he intends only his own security; and by directing that industry in such a manner as its produce may be of the greatest value, he intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention.

What Smith states is that although individuals may act solely in self-interest, a certain order, which he names *the invisible hand*, is emerging. There is, thus, a possibly positive effect which can be observed at a global level, i.e. the society, and cannot be deduced by observing the micro level, that is each individual.

Emergence is very commonly found in discussions of systems theory which take an anti-reductionism and holistic stance. Reductionism is the belief that a system's behaviour can be completely predicted and understood by analysing and observing its components; anti-reductionism being the exact opposite. Holism, being the opposite of atomism, refers to the theory that some wholes are greater than the sum of their parts or that these parts are explicable only by reference to the whole [4], a definition similar to the one given by Aristotle [2].

Although these definitions are sufficient for providing a basic understanding of emergence as a concept, they are mostly inadequate for approaching emergence as a scientific or engineering subject. Recent years have brought a renewed research interest on emergence, mostly from the communities of complex and multi-agent systems. However, to this day, a universally accepted definition of emergence is still lacking. The following section presents various attempts to define emergence and to tackle the intricate topics inherently related to it such as *novelty*, *models*, *levels*, *surprise*, and *observation*.

3 Scientific Attempts to Define Emergence

Holland famously wrote that it is unlikely that a topic as complicated as emergence will submit meekly to a concise definition [5]. While this, to this day, is true, one can gain great insight into emergence by reviewing the various definitions available in the literature, spanning different decades, sciences and disciplines.

Lewes, in 1875, proposed the use of the term *emergent* as an antonym to *resultant*. While a resultant is either the sum or the difference of two or more co-operant forces, an emergent encompasses any result which does not satisfy the aforementioned property [6]. Anderson, nearly a hundred years afterwards, managed to renew scientific interest in emergence with his claim that the whole is not only more than the sum of its parts (the original notion of emergence according to Aristotle [2]), but it is *different* [7]. Drawing a multitude of examples from molecular biology to psychology and social sciences, Anderson describes how global properties can emerge that are not deducible from the properties and behaviours of system's local components [7].

Rosen described emergence in terms of models, defining a system's behaviour as emergent if it cannot be explained anymore by the model which described the system so far [8]. This idea was later named by Cariani as "emergence relative to a model" [9], describing how any functional discrepancies between the system's model and the actual behaviour during execution can be classified as emergent.

Heylighen also states that emergence is a process which cannot be captured by a *fixed model* of a system [10]. Following a state-based approach, Heylighen uses the term model as a construct capable of capturing all possible states of a system as well as the relational constraints which define which states can be reached under different conditions. For example, the movement of a car can be typically modelled with a dynamic state space dimensionality of five: three variables to model its kinematics state, that is its coordinates in the threedimensional space, one variable for its forward velocity and another one for its yaw rate (accounting for steering). These five variables are sufficient to determine any possible state of the car model given that the dimensions of the state space are bound to be constant. If, on the other hand, the car was somehow to be split into two parts, the state space dimensionality would be immediately doubled. The original model thus would become insufficient for modelling the system after the event of splitting.

A more realistic example is crystallization where the dissolved molecules move independently of each other, causing their state space to be the product of the state spaces of all molecules, which could be virtually infinite [10]. In this scenario the initial model of the system is deemed irrelevant after an emergent phenomenon occurs (i.e. the crystallization). Heylighen proposes the use of a meta-model, which also consists of states and transition rules, capable of capturing appropriately systems which exhibit emergence. In this meta-model, each state corresponds to a simple model of the system for a given time instance during its evolution. For example one state could correspond to a model describing the car before the splitting event and another state to a model describing it after that. Transition rules in the meta-model determine the shift from one model to another.

Holland proposes a subtle change in the classical reductionism approach by contradicting the notion "that all phenomena in the universe are reducible to the laws of physics" with "all phenomena are *constrained* by the laws of physics" [5]. The latter alternative allows for the recognition of emergent phenomena even if these are the result of a relatively small set of components, governed by simple rules. Holland further claims that emergence must be the result of self-organisation and not of a centralised coordination and control [5]; an opinion Corning [11] seems to reject emphatically by arguing that self-organisation has become an academic buzzword which is often used uncritically.

Bar-Yam takes an anti-reductionism stance as well but he makes another important realisation: although it is impossible to study emergent phenomena by isolating a systems' parts and studying them separately, on their own (i.e. anti-reductionism), it is possible to study them by studying each part under the context of the system as a whole [12]. Supporting his statement he provides an example drawn from the field of neural networks:

...there are synapses between each neuron and every other neuron. If we remove a small part of the network and look at its properties, then the number of synapses that a neuron is left with in this small part is only a small fraction of the number of synapses it started with. If there are more than a few patterns stored, then when we cut out the small part of the network it loses the ability to remember any of the patterns, even the part which would be represented by the neurons contained in this part.

The important distinction BarYam makes is that the emergent behaviour is not absent in the behaviour of the individual parts (i.e. classic anti-reductionism) but rather it is *not readily observable*. If, however, the individual parts are to be studied within the context they are found, the collective behaviour can then be observed as a part of the individual components [12].

Goldstein defines emergence as "the arising of novel and coherent structures, patterns, and properties during the process of self-organisation in complex systems" [13] while he emphasises the distinction between the macroscopic level (in which the emergent phenomena are conceptualised) and the microscopic level (from whose components such phenomena arise). Goldstein attributes the following properties to any emergent phenomenon [13]:

 radical novelty: the emergent phenomenon should be novel regarding to the system and its parts and not deducible or reducible to the system's components.

- *coherence or correlation:* emergents tend to appear as integrated wholes (at the macro level).
- global or macro level: the emergent phenomena should be observable at a global or macroscopic level as opposed to the local or microscopic level in which the components lie.
- *dynamical:* emergents cannot be characterised as static properties of a system but rather they evolve, dynamically, as the system evolves over time.
- ostensive: emergents should be perceivable.

The latter point emphasises the need for an observer which plays a central role in emergence and has been the subject of great debates among researchers. Ronald, Sipper, and Capcarrere [14] are consistent with Goldstein's view of the need for an observer and taking the notion a step further they devised a scheme for testing whether a system is actually exhibiting emergent phenomena or not. The scheme requires a *designer* who designs a system in a language L_1 and an *observer*, fully aware of the design, who describes the observed macro-level system behaviour in a different language L_2 . The condition for classifying the system as capable of exhibiting emergence is the element of *surprise*, that is, the observer should be surprised by the global behaviour, documented in L_2 , denoting thus the existence of *non-obvious* causal links to the system's design in L_1 . Far from formal, the definition of "design, observation, surprise!" has been the subject of debates, especially regarding the element of surprise, how this could be formally defined and whether once the observer has gained enough insight into the causal relationships of the system, as to cease the element of surprise, the phenomenon can still be classified as emergent or not [5, 15-20].

Ronald *et al* definition is in line with Goldstein's requirement for an observer as expressed by his *ostensiveness* requirement [13]. Crutchfield shares this view regarding the necessity for an observer by claiming that the process of detecting emergence is subjective and is restricted to the observer's computational resources [21]. As an example, Crutchfield discusses the case of a self-avoidance random walking algorithm and how similarity patterns can arise from the paths traced by its execution, although the path forming process involves a lot of stochasticity. Crutchfield claims that the emergence in this scenario is in the eye of the observer whose initial predictions of the system might have failed [21].

Corning, for that matter, disqualifies the need of an observer and he claims that an emergent phenomenon is real and measurable regardless of whether someone is observing it or not [11]. Fromm shares Corning's opinion and moving a step further, he rejects both the need for the element of surprise and novelty as prerequisites for defining emergence [15]. Kubik opposes to the surprise element as well by arguing that it can be misleading and obscure better explanations [16]. Nicely put, "the moment of surprise can fade away once sufficient information is provided" and thus he suggests, towards an attempt to provide a more formal definition of emergence, to "ignore surprise" altogether [16]. Holland also believes that surprise is not an essential factor for classifying a behaviour as emergent [5] as does Damper [17]: What might be surprising on first acquaintance or at a particular stage of scientific knowledge tends to become commonplace, trite or predictable after intensive, lengthy study.

A more formal substitute for the surprise characteristic is the notion of the gap of complexity provided by Deguet and Demazeau [18]. The authors agree with Ronald's distinction (and therefore the need) of a designer and an observer but following Holland's paradigm of emergence, "much from little" [5], they devise a complexity metric which is used to prove whether a system exhibits this kind of emergent behaviour. More specifically, they consider "little" as a simple system, S, "much" as a complex phenomenon ϕ that this system exhibits and they classify the feature as emergent if the following equation results in a positive outcome:

 $eFeat = C(\phi) - C(S) \ge 0$

In the formula above, C is an abstract complexity measure. As the same complexity metric cannot be applied to both a phenomenon (ϕ) and a system (S), the authors associate a computational task to each of them [18]. The authors, however, omit to provide any practical example of a system to which they apply their proposed formula. Kubik attempts to provide a formal definition of emergence as well but firstly he suggests an informal definition of *basic* emergence [16]:

By basic emergence, we mean behavior reducible to agent-to-agent interactions without any evolutionary processes involved (i.e., an agent's behavioral set stays the same during the modeling and the analysis of the system). The environment has no rules of behavior and is changed only by the actions of agents.

Subsequently, he proceeds to formalising this notion by modelling the various concepts involved (such as a MAS, an agent, agents' behaviour) with formal constructs (grammars) and testing the proposed formal framework with the well studied emergence example of gliders. Gordon opposes in principle the need for an external observer by making a more philosophical point that if emergence requires an observer, "A relation is implied that humans are to artificial life as God is to real life. This precludes a unified treatment of life and artificial life." [19]. Instead, he suggests that both the designer and the observer should be an integral part of the system under simulation, they should emerge as well, and their emergence should also be simulated.

4 Emergence from an Engineering Perspective

Having presented the most notable attempts at defining emergence it becomes apparent that the scientific community has not yet reached a consensus on the subject. The debates seem endless regarding various issues intrinsic to emergence, such as observers, novelty, surprise, downward causation, and supervenience. Although a more thorough understanding of emergent phenomena and a more disciplined way of studying them and exploring them might shed light on some of these issues, as far as engineering emergence is concerned many of the debates might be irrelevant.

Before proceeding, it is essential to define the term engineering emergence as the word "engineering" carries strong and varying connotations in the English language. For the scope of this work, engineering emergence is used to denote a disciplined approach to introduce macroscopic behaviours by solely focusing on the microscopic level; that is to emerge out of the interactions among individuals or among those and the environment.

Consequently, the existence of *different levels* where actions of a lower level result in emergent properties in a higher one, often in *non obvious* ways, is of great importance to the engineering of emergence. The idea of levels, scales, or hierarchies is a central element in discussions about emergence which typically involve two different levels (microscopic and macroscopic) but it might be the case that intermediate levels are introduced as well (e.g. the mesoscopic [22]).

On the other hand, the element of *surprise* might lead to misconceptions, introduce ambiguity, and hinder the validity of any potential scientific results; which partly explains why there is all this controversy on the subject [5, 15– 20]. The very notion of surprise seems to contradict engineering or at the very least render any engineering attempt which might lead to surprises as erroneous. *Novelty*, however, is a far more rigid requirement for classifying a phenomenon as emergent which in turn might even surprise an observer but even if it does not, the phenomenon does not cease to be novel with regards to the system or its parts [13]. From an engineering perspective, novelty seems a reasonable requirement as engineering something that was already part of the system would have limited value while achieving a macroscopic property, such as availability, which the system or its parts did not possess beforehand, could greatly benefit the system overall.

At a first glance, *downward causation* and *supervenience* seem irrelevant to engineering emergence as they focus mostly around the philosophy of science. While downward causation is acknowledged by many researchers as a necessary ingredient for the appearance of emergent phenomena, it remains of little practical value on the quest of engineering emergence. Similarly, supervenience as a characteristic often related to strong emergence, is of limited use in engineering methodologies.

Having a first understanding of the literature on emergence and an initial view on what could assist on engineering it, the definition provided by Wolf and Holvoet [1] fits best the scope of this work and sufficiently meets the objectives already set:

A system exhibits emergence when there are coherent emergents at the macro-level that dynamically arise from the interactions between the parts at the micro-level. Such emergents are novel w.r.t. the individual parts of the system. [...] the definition uses the concept of an 'emergent' as a general term to denote the result of the process of emergence: properties, behaviour, structure, patterns, etc.

The definition incorporates the need for different levels (or points of view of the system) and the interaction between micro-level components as well as the concept of novelty. Moreover, it refers to coherence (similar to Goldstein [13]) which is essential to any dependable engineered system. At the same time it lacks any reference to the controversial issues of observers, the need for surprise, downward causation, or supervenience.

5 Discussion

Engineering emergence could greatly assist with the management of complexity, inherent to many modern digital systems. By borrowing inspiration from the nature, it should be possible to reproduce processes into artificial systems which mimic natural ones in order to allow for the emergence of complex, macroscopic behaviours such as self-adaptability, self-healing and other self-* properties.

Defining emergence is a topic of great debate by itself, with a multitude of proposed definitions in the literature spanning different disciplines and research fields. The term engineering emergence has received criticism due to popular views associating emergent phenomena with surprise, unexpectedness, and instability. This paper presented the most notable attempts to defining emergence and discussed various topics related to it, from an engineering perspective.

It seems that topics such as downward causation and supervenience are not as relevant to engineering emergence as the separation of different levels and scales of the system. Moreover, it has been argued that novelty and coherence are far more appropriate requirement for engineering systems with emergent behaviour than is the notion of surprise or unexpectedness. Finally, the definition proposed by Wolf and Holvoet [1] was adopted as a basis for future work on engineering systems with emergent properties.

The main research questions behind this work are whether emergence can be engineered, if so, to which extent, and how this could benefit digital systems. Previous work in the field has led to the development of an experimental, iterative, framework for harnessing emergent properties in artificial distributed systems [23]. The basic principle behind this framework is the separation of "normal" from emergent behaviour. While the former refers to properties and functionality which individual nodes possess in order to satisfy functional requirements, the latter represents the designer's hypotheses (for the current iteration) as to what could lead to the desired emergent phenomena. Ongoing work attempts to harness emergent properties in a generic distributed systems paradigm (named the Emergent Distributed Bio-Organization [24, 25], by introducing new properties and behaviours in the individual nodes and the environment. Future work will attempt to apply these principles on concrete case studies of the same application domain (such as decentralised Web services) or different ones (such as biology or social sciences).

References

- 1. Wolf, T.D., Holvoet, T.: Emergence versus self-organisation: Different concepts but promising when combined. In: Engineering Self-Organising Systems. (2004) 1-15
- 2. Aristotle: Metaphysics. Edited and translated by W. D. Ross. NuVision Publications (2009)
- 3. Smith, A.: An Inquiry into the Nature and Causes of the Wealth of Nations. General Books LLC (2010)
- Soanes, C., Stevenson, A.: Concise Oxford English Dictionary. 11 edn. Oxford University Press (2004)
- 5. Holland, J.H.: Emergence: From Chaos to Order. Addison-Wesley (1998)
- 6. Lewes, G.H.: The Problems of Life and Mind. Truebner (1875–1879)
- 7. Anderson, P.W.: More is different. Science 177(4047) (1972) 393-396
- 8. Rosen, R.: Anticipatory Systems: Philosophical, Mathematical, and Methodological Foundations. Oxford Press (1985)
- 9. Cariani, P.: On the Design of Devices with Emergent Semantic Functions. PhD thesis, State University of New York, Binghamton (1989)
- Heylighen, F.: Modelling emergence. World Futures: the Journal of General Evolution 31(Special Issue on Emergence) (1991) 89–104
- 11. Corning, P.A.: The re-emergence of "emergence": A venerable concept in search of a theory. Complexity **7**(6) (2002) 18–30
- Bar-Yam, Y.: Dynamics of Complex Systems (Studies in Nonlinearity. Addison-Wesley (1997)
- 13. Goldstein, J.: Emergence as a construct: History and issues. Emergence $\mathbf{1}(1)$ (1999) 49–72
- Ronald, E., Sipper, M., Capcarrere, M.: Design, observation, surprise! A test of emergence. Artificial Life 5(3) (1999) 225–239
- Fromm, J.: Ten questions about emergence. Technical Report nlin.AO/0509049, Distributed Systems Group, Kassel University (2005)
- 16. Kubík, A.: Toward a formalization of emergence. Artificial Life $\mathbf{9}(1)$ (December 2002) 41–65
- Damper, R.I.: Editorial for the Special Issue on "Emergent Properties of Complex Systems": Emergence and levels of abstraction. International Journal of Systems Science 31(7) (2000) 811–818
- Deguet, J., Demazeau, Y.: A complexity based feature to support emergence in mas. In: Proceedings of the International Central and Eastern European Conference on Multi-Agent Systems (CEEMAS), Springer (2005) 616–619
- 19. Gordon, R.: The emergence of emergence: a critique of "design, observation, surprise!". Rivista di biologia 93(2) (2000) 349–356
- Stepney, S., Polack, F., Turner, H.: Engineering emergence. In: ICECCS '06 Proceedings of the 11th IEEE International Conference on Engineering of Complex Computer Systems. (2006) 89–97
- 21. Crutchfield, J.P.: The calculi of emergence: Computation, dynamics, and induction. Physica D: Nonlinear Phenomena **75**(1–3) (August 1994) 11–54
- Ulieru, M., Doursat, R.: Emergent engineering: a radical paradigm shift. International Journal of Autonomous and Adaptive Communications Systems 4(1) (December 2011) 39–60
- 23. Eleftherakis, G., Paunovski, O., Rousis, K., Cowling, A.J.: Harnessing emergent properties in artificial distributed networks: an experimental framework. In Ritson,

C., Andrews, P., Stepney, S., eds.: 4th Workshop on Complex Systems Modelling and Simulation, Paris, France, Luniver Press (August 2011) 141–144

- Paunovski, O., Eleftherakis, G., Rousis, K., Cowling, A.J.: Emergent distributed bio-organization: A framework for achieving emergent properties in unstructured distributed systems. In: 6th International Symposium on Intelligent Distributed Computing (IDC 2012), Calabria, Italy, Studies in Computational Intelligence, Springer (September 2012) to appear.
- Rousis, K., Eleftherakis, G., Paunovski, O., Cowling, A.J.: Formal modelling of a bio-inspired paradigm capable of exhibiting emergence. In: 5th Balkan Conference in Informatics (BCI 2012), Novi Sad, Serbia, ACM (September 2012) to appear.

Stability properties of stochastic neural networks

Biljana Tojtovska

Faculty of Computer Science and Engineering, Skopje Bulevar Rudjer Bošković b.b., 1000, Skopje biljana.tojtovska@finki.ukim.mk

Abstract. Artificial neural networks can not accurately describe complex dynamical systems which have intrinsic noise. In order to improve their performance, different stochastic modifications have been proposed. The goal of this paper is to discuss some models of stochastic neural networks and their stability properties and give a new result of stability of Cohen-Grossberg neural network with respect to a general decay function. First we describe the deterministic Hopfield model and motivate the needs for its stochastic modifications. After presenting the Boltzmann machines, we discuss the general model of a stochastic neural network with a delay, given by a system of stochastic differential equations (SDEs). Last, we discuss the Cohen-Grossberg neural network with timevarying delays and present a generalisations of a known result refering to the pth moment exponential stability. We give a numerical example which supports the theory.

Keywords: Stochastic neural networks, stochastic differential equation, stability analysis, general decay function

1 Introduction

The neural networks are mathematical models based on the concepts of parallel distributed processing, learning and self-organisation. They are inspired by the structure of biological neural networks and by the way they process information. They are decision making tools which have been successfully applied in solving optimisation problems, speech and pattern recognition problems as well as image analysis. Though they proved to be useful, they fail to give an accurate description of the dynamics of biological neural networks. Von Neumann in his famous book "The computer and the brain" [1] concludes that the brain uses its own statistical language, Taylor in [2] describes a biologically realistic model of synaptically noisy neurons and Sejnowski in [3] presents the brain as a stochastic dynamical system. Another reason for introducing stochastic elements in the network comes from practical application - input data very often come from a system with noise. Different stochastic models were introduced and they showed improved performance when applied to different problems and gave a more realistic picture of natural dynamical systems. In this paper we give a description of the famous Hopfield neural network model and then give motivation for some of its stochastic modifications - Boltzmann and diffusion machines. Then we consider the general case of stochastic neural networks which are given by a system of stochastic differential equations and we also consider the case when a delay is present in the network. This covers the stochastic modification of many known networks like the Cohen-Grossberg model, cellular neural networks and BAM networks (neural networks with bidirectional associative memory). These models are very complex and for their analysis we need the theory of stochastic processes. We discuss the stability properties of the neural networks, as an important feature of any dynamical system. This is a big research area and there are many open problems since there are no general methods that prove the stability of any model. Some authors have discussed the mean-square, pth moment and almost sure exponential stability of different models of stochastic neural networks - see [4–11] and references therein. In this paper we give conditions for the pth moment stability with respect to a general decay function which is a generalisation of a known result presented in [4].

2 Hopfield neural network model

The work done by Hopfield ([12], [13]) has made an immense impact on the study of neural networks. He has proposed a single-layered network model of n fully connected units, which is based on the Hebbian learning algorithm. The global state of the network $x(t) = (x_1(t), x_2(t), ..., x_n(t))$ at time t is described by the energy of that state E(x(t)). The energy function proposed by Hopfield for the discrete model is given by

$$E(x(t)) = -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} x_i(t) x_j(t) + \sum_{i=1}^{n} \theta_i x_i(t)$$
(2.1)

and depends on the given threshold values θ_i , output of the units x_i and the weights w_{ij} which describe the strength of the connection between the *i*th and *j*th neuron. Hopfield has shown that in an asynchronously updated network¹, that has bipolar or binary output functions, symmetric weight matrix and no self-feedback ($w_{ii} = 0$), the energy function is non-increasing with respect to the dynamics of the network. This means that with every unit update the energy of the network state does not increase. Thus the dynamic of the network leads towards energy minima, which on the other hand correspond to the network stable states. These stable states or equilibrium states are used for pattern storage in the network.

In the continuous Hopfield model the outputs of the neurons change continuously and simultaneously. The output function f(t) is continuous, non-linear, usually a sigmoid function. In [13] Hopfield has proposed the following energy function

$$E(x(t)) = -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} x_i(t) x_j(t) + \sum_{i=1}^{n} \int_0^{x_i(t)} f^{-1}(t) dt$$
(2.2)

¹ The update of the network is asynchronous when one unit at a time is chosen at random and its state is updated.

Again, when $w_{ij} = w_{ji}$, using the properties of the output function we can conclude that the energy does not increase as the network changes in time. This means that there are basins of attractions with stable points which correspond to local minima of the energy function. This makes the network useful for pattern storage.

The analysis of the energy function gives an idea how the neural networks can be used for computation of local minimum of a given function E(x) defined on a hypercube $[0,1]^n$. It has been shown that many complex computational problems can be reduced to finding the local minimum of a given quadratic function [14]. In this and similar optimisation problems, if the weights can be identified from the energy function, we can build a neural network and using its dynamics we can reach the equilibrium states. This search is based on local search in the discrete case and gradient-descent method in the continuous case. Thus the network may get stuck at a local minimum i.e. it will not be able to go at states with higher energy and find the global minimum of a function if needed. This is very inconvenient since many optimisation problems have a lot of local minima. Similar difficulty arises in the pattern storage problem. If the network has a higher storage capacity than the number of patterns we want to store, we may reach a false local minimum and recall a pattern which is not one of those that we stored. The greater the number of false minima, the greater the probability of error in recalling. Different approaches have been proposed to overcome this problem. One of them is introducing stochastic updates of the network and using the method of *simulated annealing*. This brings us to the concept of Boltzmann and diffusion machines.

3 Boltzmann machines

A Boltzmann machine [15–18] is a fully (bidirectionally) connected network of n units with symmetric connection weights w_{ij} and bias of the units θ_i . No selfconnection is allowed in the network. We consider the discrete case, where the configuration of the network at time $t, x(t) = (x_1(t), ..., x_n(t))$, is a $\{0, 1\}^n$ -valued discrete-time Markov chain. The 0-1 values here correspond to the activation state of the unit². The main difference with the described Hopfield model is the stochastic update of the network. At time t + 1 a unit i is randomly chosen³ and is updated with the following probabilities $\mathbb{P}(x_i(t+1) = 1) = p_i(t+1)$ and $\mathbb{P}(x_i(t+1) = 0) = 1 - p_i(t+1)$ where

$$p_i(t+1) = \frac{1}{1 + e^{-(\sum_{j=1}^n w_{ij} x_j(t) - \theta_i)/T}}$$
(3.1)

with T being a "temperature" parameter. This way the new configuration $x(t + 1) \in \{0, 1\}^n$ is obtained. The energy of the network is given as in the discrete Hopfield model with equation (2.1).

 $^{^{2}}$ A Boltzmann machine can be also defined with bipolar states.

 $^{^{3}}$ We take here that each unit is chosen with equal probability.

Since the network is stochastic we can not speak of stable states in which the network stays once it reaches them. In this case we can consider the probability distribution of the network configurations and analyse whether it can reach an equilibrium. Given the probabilities (3.1) we can calculate the conditional probabilities P(x'|x) for change of the network state and that way we will obtain the transition matrix M of the described Markov chain. Given an initial distribution p(0), the distribution after k updates of the network will be given by $p(0)M^k$. In such a described network with non-zero probability we can reach any state from any given state in a finite number of steps. Also, the expected number of steps needed for revisiting a given state is finite. As a result a stationary distribution π exists ($\pi = \pi M$) and $p(k) = p(0)M^k \to \pi$ when $k \to \infty$. Thus, if we run the network long enough the probability distribution of the visited states will reach an equilibrium called thermal equilibrium. In [16] it is shown that for a fixed T the stationary distribution is given by

$$\pi(x') = \frac{e^{-\frac{1}{T}E(x')}}{\sum_{x} e^{-\frac{1}{T}E(x)}} = \frac{e^{-\frac{1}{T}E(x')}}{Z}$$
(3.2)

which is actually the Boltzmann distribution. The main characteristic of this distribution is that a system in thermal equilibrium will stay in the states with lower energy, with higher probability. The system can jump to higher energy states but the probabilities of this happening decrease when $T \to 0$. Thus we obtain the Hopfield model as a limit case when $T \to 0$.

The continuous version of the model is given by the following stochastic differential equation

$$d(x_i(t)) = -\frac{\partial}{\partial x_i} E(x(t)) + \sqrt{2T} dW_i(t) \quad i = 1, 2, ..., n$$

$$(3.3)$$

where W_i are independent Brownian motions. We want to know when a stationary distribution of the form (3.2) exists. For general $x \in \mathbb{R}^n$ the answer depends on the integrability of $e^{\frac{E(x)}{T}}$. To ensure the existence of a stationary Boltzmann distribution when $x \in \{0, 1\}^n$ we have to put some boundary conditions which will ensure that the stochastic process will not exit the hypercube $\{0, 1\}^n$.

Wong in [16] gives a modification of the previously described models, obtained from the Hopfield network, by adding noise at each node. He showed that, if the dynamics of the network are given by the following system of SDEs

$$x_{i}(t) = g(u_{i}(t))$$

$$d(u_{i}(t)) = -E_{i}(x(t))dt + \sqrt{\frac{2T}{g'(u_{i}(t))}}dW_{i} \quad i = 1, .., n$$
(3.4)

where $E_i(x(x)) = \frac{\partial}{\partial x_i} E(x(t))$, then it is a Markov process with stationary Boltzmann distribution. Such a model is called a *diffusion machine*. For details and applications we refer to [16].

4 Stochastic neural networks - general models and stability analysis

Hopfield's constraint on symmetric weight connections and some additional conditions on the output functions do not apply to biological systems. Change of the constraints and even imposing random weights on the connections, makes the analysis of the network behaviour much more difficult. But, since the inputoutput data as well as the other variables of the networks can be seen as a realisation of a stochastic process, we can use the theory of stochastic differential equations to find some important properties of the stochastic model. Next we present some of the theory needed for analysis of a stochastic neural network given by a system of stochastic differential equations.⁴

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq t_0}, \mathbb{P})$ be a complete probability space where $t_0 \geq 0$ and $(\mathcal{F}_t)_{t \geq t_0}$ is the natural filtration with respect to the *m*-dimensional Brownian motion $W(t) = (W_1(t), ..., W_m(t))^T$ defined on this probability space. The filtration satisfies the usual conditions i.e. it is right-continuous and \mathcal{F}_{t_0} contains all the \mathbb{P} -null sets. For $t_0 \geq 0$ let x_0 be an \mathcal{F}_{t_0} -measurable \mathbb{R}^n -valued random variable with $\mathbb{E}(x_0^2) < \infty$.

A general model of a stochastic neural network is given by the following n-dimensional stochastic differential equation

$$dx(t) = [-Bx(t) + Ag(x(t)) + C]dt + \sigma(x(t))dW(t)$$
(4.1)

Here $x(t) = (x_1(t), ..., x_n(t))^T$, $g(x(t)) = (g_1(x_1(t)), ..., g_n(x_n(t)))$ where x_i models the voltage on the input of the *i*-th neuron and g_i is the activation function characteristic for the *i*-th neuron. $B = diag(b_1, ..., b_n)$, $A = (a_{ij})_{n \times n}$ and $C = diag(c_1, ..., c_n)$ are matrices which depend on some network characteristics and $\sigma(x(t)) = (\sigma_{ij}(x(t)))_{n \times n}$ is the diffusion matrix. More details on the architecture can be found in [20].

Since a time delay exists in the processing and transformation of signals, this delay in the communication channels can be included in the model. The models can be made more general by for example taking different types of activation functions and adding some additional properties of the network which will make the matrices A, B and C time dependent. The time delays τ_i can also be made time dependent or they can even be random variables. Thus we get the stochastic modification of many already known deterministic models such as Cohen-Grossberg model, cellular neural networks and BAM networks (neural networks with bidirectional associative memory). With this, new models are also defined and they can be practically tested.

One of the basic questions that have to be answered is when a given SDE has a unique solution. However, even when we do know that the solution exists, very often it is not explicitly given. Still, both from theoretical and practical point of view, it is important to know its characteristics. The stability property of the solution is important for any dynamical system. In the case of neural networks

⁴ We follow the theory presented in [19].

it means existence of an equilibrium state and thus is important for the process of learning and pattern formation. It is equally important to know when and how a stable network can be destabilised by addition of stochastic perturbation and vice versa. 5

When working with a stochastic neural network model (4.1), we can analyse few types of stability - stochastic stability, mean-square, pth moment and almost sure exponential stability. Many methods are already used for the stability analysis - Lyapunov stability theory, Razumikhin type techniques, method of variation of parameters, semimartingale theory, the well known inequalities of Hölder and Minkowski applied to stochastic processes, Lassale's method and other results from matrix theory. Using these methods some authors have discussed the stability characteristics of different models of stochastic neural networks (see for [4–11] and references therein). Still these methods can not be used in general which leaves many open problems. In many cases the stability with respect to a general decay function is of interest. Though some results for general decay stability of functional differential equations with delay do exist (ex. [21]), to the best of our knowledge only few of them address the models of stochastic neural networks (ex. [22]). The aim of our research is to fill in this gap. In this paper we give conditions on the *p*th moment stability with respect to a general decay function for a model of Cohen-Grossberg neural network with time varying delays, defined for $t \geq t_0$

$$dx(t) = -H(x(t))[C(t, x(t)) - A(t)F(x(t)) - B(t)G(x_t)]dt + \sigma(t, x(t), x_t)dW_t$$
(4.2)

with an initial condition $x_{t_0}(s) = x(t_0 - s) = \xi(s)$ for $s \in [t_0 - \tau, t_0]$.

Here $x(t) = (x_1(t), ..., x_n(t))^T$ is an \mathbb{R}^n -valued stochastic process describing the state variables at time t, which correspond to the n neurons of the network, τ_i are the delays in the *i*-th node, $\tau = \max_{i \in \{1,...,n\}} \tau_i, x_\tau(t) = (x_1(t-\tau_1), ..., x_n(t-\tau_n))$ is the delayed process. $\xi = \{\xi(\theta) : \theta \in [t_0 - \tau, t_0]\} \in \mathcal{L}_{\mathcal{F}_{t_0}}([t_0 - \tau, t_0]; \mathbb{R}^n)$, i.e. it is \mathcal{F}_t -adapted, $C([t_0 - \tau, t_0]; \mathbb{R}^n)$ -valued random variable.

In the model $H(x(t)) = diag(h_1(x_1(t)), ..., h_n(x_n(t)))$, where $h_i(x(t))$ are amplification functions, and $C(t, x(t)) = (c_1(t, x_1(t)), ..., c_n(t, x_n(t)))^T$, where $c_i(t, x_i(t))$ are appropriately behaved functions dependent on t and on the state process x(t). Matrices $A(t) = [a_{ij}(t)]_{n \times n}$ and $B(t) = [b_{ij}(t)]_{n \times n}$ describe the strength of the neuron interconnections in the network at times t and $t - \tau_j(t)$, respectively, while $F(x(t)) = (f_1(x_1(t)), ..., f_n(x_n(t)))^T$ and $G(x_t) = (g_1(x_{1,t}), ..., g_n(x_{n,t}))^T$ are vectors of activation functions, where f_j and g_j denote the output of the *j*th unit at times t and $t - \tau_j(t)$, respectively. The term $\sigma(t, x(t), x_t) = [\sigma_{ij}(t, x_j(t), x_{j,t})]_{n \times n}$ is a diffusion-coefficient matrix⁶.

⁵ Definitions and theoretical results on these topics can be found in [19].

⁶ For details of the network parameters we refer to [18] or [23] where the deterministic model is discussed.

This model is generalisation of the ones used in [4] and [5], where the authors prove pth moment and almost sure exponential stability. We generalise the model by adding time delay in the diffusion matrix and we generalise the drift coefficient. We further extend their work by analysing stability with respect to a general decay function, which includes the special case of exponential stability. We give here the following definition of stability with respect to a general decay function.

Definition 1. Let $\lambda(t) \in C([\min\{\tau, t_0 - \tau\}, +\infty); \mathbb{R}^+)$ be a decay function which is strictly increasing, tends to infinity when $t \to \infty$ and satisfies $\lambda(t) \ge \lambda'(t)$ and $\lambda(s+t) \le \lambda(s)\lambda(t)$ for all s,t in its domain. We say that the trivial solution of equation (4.2) is **pth moment stable with a decay** $\lambda(t)$ and of **order** γ if there are constants $\gamma > 0$ and $c = c(\xi) > 0$ s.t.

$$\mathbb{E}||x(t,\xi)||^p \le c(\xi)\lambda^{-\gamma}(t), \qquad t \ge t_0$$

holds for any $\xi \in \mathcal{L}^p_{\mathcal{F}_{t_0}}([t_0 - \tau, t_0]; \mathbb{R}^n)$. The trivial solution of (4.2) is said to be almost surely stable with decay $\lambda(t)$ and of order γ if

$$\limsup_{t \to \infty} \frac{\ln||x(t;\xi)||}{\ln\lambda(t)} \le -\gamma \quad a.s.$$

We assume that the parameters of the equation (4.2) satisfy the general conditions needed for existence and uniqueness of the global solution $x(t;\xi)$ on $[t_0, +\infty)$. Since we discuss *p*th moment stability ($p \ge 2$), in order to have $\mathbb{E} \sup_{t \in [t_0, +\infty)} ||x(t,\xi)||^p < \infty$ we also need to assume that $\xi \in \mathcal{L}^p_{\mathcal{F}_{t_0}}([t_0 - \tau, t_0]; \mathbb{R}^n)$. Before stating our main result, we give further assumptions for the functions in

the model (4.2). For i, j = 1, 2, ..., n, we suppose that:

 $(\mathbf{H_1})$ There exist positive constants $\underline{h}_i, \overline{h}_i$ such that

$$0 < \underline{h}_i \le h_i(x) \le \overline{h}_i, \quad x \in \mathbb{R}.$$

(**H**₂) There exists a positive function $\alpha_i(t), t \in [t_0, \infty)$ such that

$$xc_i(t,x) \ge \alpha_i(t)x^2$$
 $(t,x) \in [t_0,\infty) \times \mathbb{R}.$

(**H**₃) The activation functions are Lipshitz continuous, that is, there exist positive constants β_i, δ_i such that

$$|f_i(x) - f_i(y)| \le \beta_i |x - y|, \quad |g_i(x) - g_i(y)| \le \delta_i |x - y|, \quad x, y \in \mathbb{R}.$$

(**H**₄) There exist nonnegative functions $\hat{a}_{ij}(t)$ and $\hat{b}_{ij}(t), t \in [t_0, \infty)$ such that

$$\max\{|a_{ij}(t)|, |a_{ji}(t)|\} \le \hat{a}_{ij}(t), \ |b_{ij}(t)| \le b_{ij}(t), \ t \in [t_0, \infty).$$

(**H**₅) The drift functions $\sigma_{ij}(t, x, y)$ are globally Lipshitz continuous, uniformly in t, and there exist nonnegative constants μ_{ij} and nonnegative functions $\nu_{ij}(t), t \in [t_0, \infty)$ such that

$$\sigma_{ij}^2(t, x, y) \le \mu_{ij} x^2 + \nu_{ij}(t) y^2, \quad (t, x, y) \in [t_0, \infty) \times \mathbb{R}^2.$$

First we present a lemma which will help us state the main result.

Lemma 1. Let $\delta > 0$ be a given constant and a(t), $b(t) \in C([t_0, \infty); \mathbb{R}^n)$ be positive functions. Let $\delta \ge 1$ and there exist constants $a_0 > 0$ and $0 < \eta < 1$ such that $a(t) \ge a_0$ and $b(t) \le \eta a(t)$ for $t \ge t_0$. Then, for any $t \ge t_0$ there exists a unique positive solution $\gamma = \gamma(t)$, for the equation

$$\gamma = a(t) - b(t)\delta^{\gamma}. \tag{4.3}$$

In addition $\gamma^* := \inf_{t \ge t_0} \gamma(t) > 0.$

Next we state the main result concerning the pth moment stability with respect to a general decay function for the Eq. (4.2).

Theorem 1. Let $\xi \in \mathcal{L}^p_{\mathcal{F}_{t_0}}([t_0 - \tau, t_0]; \mathbb{R}^n)$ and let the assumptions $(\mathbf{H_1}) - (\mathbf{H_5})$ hold for Eq. (4.2). Also, let functions a(t), b(t), $t \in [t_0, \infty)$ be defined as

$$a(t) = \min_{i} \left\{ p \underline{h}_{i} \alpha_{i}(t) - \sum_{j=1}^{n} \left[p \hat{a}_{ij}(t) \left(\overline{h}_{i} \beta_{j} \right)^{\frac{p-1}{p}} \left(\overline{h}_{j} \beta_{i} \frac{m_{j}}{m_{i}} \right)^{\frac{1}{p}} + (p-1) \overline{h}_{i} \delta_{j} \hat{b}_{ij}(t) + \frac{p(p-1)}{2} \mu_{ij}^{\frac{p-2}{p}} \left(\mu_{ji} \frac{m_{j}}{m_{i}} \right)^{\frac{2}{p}} + \frac{(p-1)(p-2)}{2} \nu_{ij}(t) \right] \right\} > 0, \quad (4.4)$$

$$b(t) = n \max_{i} \left\{ \sum_{j=1}^{n} \frac{m_j}{m_i} \left[\overline{h}_j \delta_i \, \hat{b}_{ji}(t) + (p-1)\nu_{ij}(t) \right] \right\},\tag{4.5}$$

where $p \ge 2$ and $m_i, i = 1, ..., n$ are some positive constants. Let $\lambda(t)$ be a decay function. If a(t) and b(t) satisfy the conditions from Lemma 1 with $\delta = \lambda(\tau)$, then Eq. (4.2) is pth moment stable with a decay function $\lambda(t)$ of order γ^* .

The results in [4] and [5] are a special case of Theorem 1. To the best of our knowledge this result is new and has not been published before. ⁷ Next we give a numerical example which supports our theory.

5 Numerical example

Let us consider the 4th moment general decay stability of the stochastic Cohen-Grossberg neural network with time-varying delays, represented by Eq. (4.2),

⁷ The results are submitted for a review (B.Tojtovska, S.Jankovic, *Decay stability of stochastic Cohen-Grossberg neural networks with time-varying delays*).

where $x_0 = \xi \in \mathcal{L}^4_{\mathcal{F}_0}([-\tau, 0]; \mathbb{R}^2)$, W(t) is a two-dimensional Brownian motion, $\tau_1(t) = \tau_2(t) = |sin(t)|$ are delay functions, $x = (x_1, x_2)^T, y = (y_1, y_2)^T \in \mathbb{R}^2$, $t \ge 0$ and

$$\begin{split} H(x) &= \begin{bmatrix} 1.6 + 0.1 \sin(x_1(t)) & 0 \\ 0 & 1.6 + 0.1 \cos(x_2(t)) \end{bmatrix}, \\ C(t,x) &= \begin{bmatrix} (11+5t)x_1(t), (10.5+5t)x_2(t) \end{bmatrix}^T, \\ A(t) &= \begin{bmatrix} 0.4 + 0.1t & 0.2 + 0.1t \\ 0.2 + 0.1t & 0.2 + 0.1t \end{bmatrix}, \quad B(t) = \begin{bmatrix} 0.02t & 0.02 + 0.01t \\ 0.01 & 0.02 + 0.01t \end{bmatrix}, \\ f(x) &= \begin{bmatrix} 4 \tanh(x_1), 4 \tanh(x_2) \end{bmatrix}^T, \quad g(y) = \begin{bmatrix} 4 \tanh(y_1), 4 \tanh(y_2) \end{bmatrix}^T, \\ \sigma(t,x,y) &= \begin{bmatrix} 3.12^{\frac{1}{4}} |x_1y_1|^{\frac{1}{2}} & 1.76^{\frac{1}{4}} |x_2y_2|^{\frac{1}{2}} \\ 3.12^{\frac{1}{4}} |x_1y_1|^{\frac{1}{2}} & 1.76^{\frac{1}{4}} |x_2y_2|^{\frac{1}{2}} \end{bmatrix}. \end{split}$$

Let p=4 and let us consider the polynomial function $\lambda_1(t) = t + 1$. Given the parameters, we can compute that a(t) = 26.882 + 23.948t, b(t) = 12.544 + 0.272t and $\lambda_1(\tau) = \lambda_1(1) = 2$. The functions a(t) and b(t) satisfy the condition in Lemma 1 and thus the equation

$$\gamma = 26.882 + 23.948t - (12.544 + 0.272t)2^{\gamma} \tag{5.1}$$

for any $t \ge 1$ has a unique solution $\gamma = \gamma(t)$. Fig. 1. shows the plot of the function $\gamma(t)$. Obviously $\gamma^* = \gamma(1) \approx 1.93183$.

For the presented model all the assumptions (H1)-(H5) hold. We can conclude that the trivial solution is 4th moment stable with rate $\lambda_1(t) = t + 1$ and order $\gamma^* = 1.93183$. Similarly, when $\lambda_2(t) = \exp(t)$ and $\lambda_2(\tau) = e$ we can conclude that the trivial solution is 4th moment exponentially stable of order $\gamma_2^* = 1.35086$.

It can be also shown that the solution is almost sure stable with respect to the decay functions $\lambda_1(t)$ and $\lambda_2(t)$ and the corresponding rates can be computed. Fig.2 represents the 4th moment of the process and Fig.3 shows numerical simulations of the processes $x_1(t), x_2(t)$ for the given example.

6 Conclusions

In this paper we give a motivation for analysis of stochastic neural network models. After introducing the Hopfield model we discuss the Boltzmann and diffusion machines and see the importance of the stability property of the network. Next we consider general models given by a system of stochastic differential equations. We give a new result for pth moment stability with respect to a general decay function of a general Cohen-Grossberg model. At the end we give a numerical example and a simulation of the stochastic processes which additionally supports our theory.

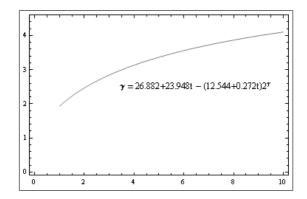


Fig.1. Function $\gamma = \gamma(t), \ \gamma^* = 1.93183.$

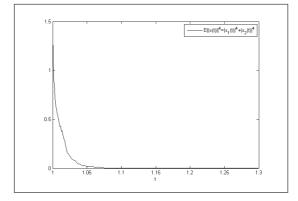


Fig.2. 4th moment $\mathbb{E}||x(t)||^4$ for the given example.

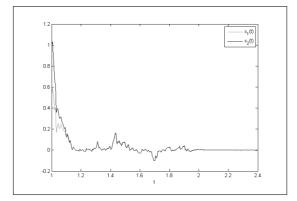


Fig.3. Simulation of the processes $x_1(t)$ and $x_2(t)$.

References

1. von Neumann, J. : The computer and the brain. New Haven: Yale University Press , 1958

- Taylor, J.G.: Spontaneous behaviour in neural networks. J. Theor. Biol.(1972), Vol. 36, Issue 3, 513-528, 1972
- Sejnowski, T.J.: Skeleton filters in the brain. in Hinton, G.E., Anderson, J.A. (Eds.): Parallel Models of Associative Memory (pp 189-208), Hillsdale LEE, New Jersey, 1981
- 4. Huang, C., Cao, J. : Stochastic dynamics of nonautonomous Grossberg neural networks. Abstract and Appl. Anal., Vol 2011, Article ID 297147
- Huang, C., He, Y., Chen, P. : Dynamic analysis of stochastic recurrent neural networks. Neural Process. Lett., 27 (2008), 267-276
- Liao, X.X., Mao, X. : Stability of stochastic neural networks. Neural, Parallel and Sci. Computations, 4(1996), 205-224
- Liao, X.X., Mao, X. :Exponential stability and instability of stochastic neural networks, Stoch. Anal. Appl., 14 (2) (1996), 165-185.
- 8. Blythe, S., Mao, X., Liao, X. : Stability of stochastic delay neural networks, J. of the Franklin Institute, 338 (4) (2001), 481-495.
- 9. Blythe, S., Mao, X., Shah, A. : Razumikhin-type theorems on stability of stochastic neural networks with delays, Stoch. Anal. Appl. 19 (1) (2001), 85-101.
- Zhou, Q., Wan, L. : Exponential stability of stochastic delayed Hopfield neural networks, Appl. Math. Computation 199 (2008), 84-89.
- Sun, Y., Cao, J. : pth moment exponential stability of stochastic recurrent neural networks with time-varying delays, Nonlinear Analysis: Real World Appl. 8 (4) (2007), 1171-1185.
- Hopfield, J.J. : Neural networks and physical systems with emergent collective computational abilities. Proc. NatL Acad. Sci. USA, April 1982, Vol. 79, Biophysics, pp. 2554-2558
- Hopfield, J.J. : Neurons with graded response have collective computational properties like those of two-sate neurons. Proc. Nat. Acad. Sci. USA (1984), Vol. 81, Biophysics, pp. 3088-3092
- Hopfield, J.J., Tank, D.W. : Neural computation of decisions optimization problems. Biol.Cybernet, 52(1985), 141-152
- 15. Rojas, R.: Neural networks A systematic Introduction. Springer-Verlag,Berlin 1996
- 16. Wong, E. : Stochastic neural networks. Algorithmica (1991) 6, 466-478
- 17. Ackley, D.H., Hinton, G.W., Sejnowski, T.J. : A learning algorithm for Boltzmann machines. Cognitive Sci. 9 (1985), 147-169
- 18. Yegnanarayana, B. : Artificial neural networks. PHI Learning Pvt. Ltd., 2004
- 19. Mao, X.: Stochastic differential equations and applications. 2nd Ed., Woodhead Publishing, 2010
- Hu, S., Liao, X.X., Mao, X. : Stochastic Hopfield neural networks. Journal of Physics A: Mathematical and General 36 (2003), 22352249
- Pavlović, G., Janković, S. : Razumikhin-type theorems on general decay stability of stochastic functional differential equations with infinite delay, J. Comp. Appl. Math. 236 (2012), 1679-1690.
- 22. Hu, Y., Huang, C. : Lasalle method and general decay stability of stochastic neural networks with mixed delays, J. Appl. Math.Computing 38 (1-2) (2012), 257-278.
- 23. Cohen, M.A., Grossberg, S. : Absolute stability of global pattern formation and parallel memory storage by competitive neural networks, IEEE Transactions on Systems, Man. and Cybernetics, 13 (5) (1983), 815-826.

Tools for Visual Simulation of MAS Models

Isidora Petreska

South East European Research Centre (SEERC), 24 Proxenou Koromila Str., Thessaloniki 54622, Greece, ispetreska@seerc.org

Abstract. Modelling and verification of spatial multi-agent systems present many interesting challenges, especially when it comes to reliability and robustness. Among other characteristics, these systems are commonly composed of spatial properties which typically cannot be formally verified due to combinatorial explosion. Verification of these systems can be supported by observing the visual outcome of executable simulation scenarios and comparing it with the expected behaviour of the system. In some cases this can lead to identifying new patterns (emergence). This short paper reviews three different simulation platforms, namely NetLogo, Repast and FLAME, under various performance factors targeted to spatial state-based systems.

Keywords: Simulation, spatial MAS, NetLogo, Repast, FLAME

1 Introduction

Spatial agents can be defined as collections of agents distributed throughout a physical space and possessing movement capabilities. Additionally, agents might have incomplete knowledge of the environment. Formal verification and validation are not always feasible to be achieved in these systems. This constraint rises from the fact that some of the properties found in these systems can have *infinitely many* values, which in turn leads to state space explosion. Moreover, all of the system's properties should be identified first, before there is an attempt to be verified, properties and behaviour related to the positioning in space (or emergent behaviour such as line formation, flocks, schools, herds etc.) may occur in a system even though they were not explicitly modelled. This leads to the fact that spatial MAS may exhibit behaviour that was not present in original system's requirements. In order to detect systems that have such characteristics, simulation tools providing visual output should be applied.

There are many large-scale state-based MAS simulation platforms [1], such as Repast [2], Swarm [3], NetLogo [4], MASON [5], and FLAME [6], to name a few. Such platforms may facilitate the informal verification of a spatial model by allowing researchers to compare the simulation's outcome with the expected behaviour of the system or to discover new patterns (emergence). This short paper presents a survey of NetLogo, Repast and FLAME, in regards to simulating MAS with spatial characteristics. The review is based on comparing how well they perform in modelling, implementation, verification, testing, visualisation and other factors.

Section 2 provides an overview of NetLogo, Repast and FLAME, supported by a case study and code for these visual simulations platforms. A critical comparison is presented in Section 3. Finally, all the results are summarized in Section 4.

2 Overview of NetLogo, Repast and FLAME

NetLogo [4] is a simulation platform for visual animation of multi-agent systems. It is supported by a functional language that could represent agent's behaviour, as well as an environment for creation of a graphical user interface. As a programmable modelling environment, NetLogo is considered specialised into simulating natural and social phenomena, including modelling of complex systems [7].

Repast (REcursive Porous Agent Simulation Toolkit) [2] is a framework for creating agent-based simulations. This platform allows specification of logical and spatial structure of a model, what types of agents a model is composed of, and the individual properties and behaviour of the agents themselves. In this work we are interested in its latest release *Repast Simphony*.

FLAME (Flexible Large-scale Agent Modelling Environment) is an agent based modelling framework that employs X-machines [8] as the basic computational model [6]. FLAME is specialised to support mainly projects related to cell biology, including tissue cultures and signalling pathways. The FLAME models can be described with XMML, X-machine mark-up modelling language used for defining agents and their communication links. The functions are implemented in the C programming language.

The following case study, known as the aggressor-defender game [9], was modelled in each of these platforms. In this example, there are two teams of agents randomly distributed in an environment: defenders (or *friends*) and aggressors (or *enemies*). There are two different sub-games involved:

- All the agents defend at each turn everyone tends to position between a friend and an enemy (such as they were defending the friend against the enemy), Fig. 1 a).
- All the agents flee at each turn everyone tries to position in a way that a friend is between themselves and an enemy (such as the friend protects them from an enemy), Fig. 1 b).

The visual animation output for the aggressor-defender game, shows an observable emergent spatial behaviour which clearly demonstrates that suitable simulation tool for modelling spatial MAS can help into detection of unknown spatial behaviour:

- The model in which all the agents defend, see Fig. 1 a), behaved as all the agents quickly collapsed into a tight knot.
- The model in which all the agents flee, see Fig. 1 b), behaved as a highly dynamic group that expands over time towards the ends of the environment.

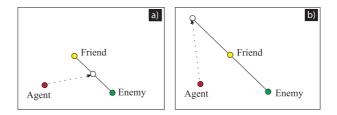


Fig. 1. Rules for playing the aggressor-defender game.

3 Comparison of the Simulation Platforms

This section provides a comparison between NetLogo, Repast, and FLAME, based on the following factors: modelling, design, implementation, verification, validation, testing, visualisation, supported size of MAS and high performance computing. Table 1 demonstrates the function *defend* modelled in each of these platforms.

Table 1. The function defend in NetLogo, Repast and FLAME for the aggressor-defender game

	FUNCTION defend
NetLogo	<pre>to defend facexy ([xcor] of friend + [xcor] of enemy) / 2 ([ycor] of friend + [ycor] of enemy) / 2 end</pre>
Repast	<pre>def defend() { facexy(({ xcor }.of(friend) + { xcor }.of(enemy)) / 2, ({ ycor }.of(friend) + { ycor }.of(enemy)) / 2)}</pre>
FLAME	<pre>double handle_defend_X(double friendX, double enemyX) { double newPosition = (friendX + enemyX)/2; return newPosition;} double handle_defend_Y(double friendY, double enemyY) { double newPosition = (friendY + enemyY)/2; return newPosition;}</pre>

3.1 Modelling and Design

Observing the snippets of code provided in Table 1, it becomes apparent that NetLogo and Repast have a very similar syntax; which sets them apart from FLAME. The NetLogo programming language is fairly simple, a procedural language based on follow up languages of Logo. The Repast Simphony model was build using ReLogo which is a dialect of Logo. However, Repast Simphony models can also be built by utilizing point-and-click flowcharts. When it comes to defining agents and their communication links, NetLogo and Repast lack of modelling methodology. Instead, they provide an implementation language with which the user is expected to model the desired system. FLAME uses XMML which enforces the modeller to think in terms of design. This represents a clear advantage of this framework.

3.2 Implementation

Movement functions and manipulation of an agent's position and direction, are the most important spatial characteristics. Unlike FLAME, NetLogo and Repast support these characteristics natively. As it can be seen in Table 1, the FLAME implementation of the coordinates of an agent are a custom user variable of any type chosen by the modeller. This notion makes it more difficult to write FLAME models for spatial MAS or to understand existing models. Furthermore, there is an increased level of difficulty when it comes to writing movement functions in FLAME. In NetLogo and Repast this is acheived with the built-in function facexy.

NetLogo and Repast have the advantage of assigning agents to positions randomly, therefore allowing for the set up of different simulation scenarios automatically. FLAME requires that agents are individually instantiated in a set up file. This usually involves writing an external program which will help into generation of data for the set up file.

3.3 Verification, Validation and Testing

NetLogo and Repast do not support verification or testing techniques [4, 10]. There is, however, active research towards testing techniques for Repast, which includes a generic testing framework for agent-based simulation models [11, 10]. Finally, given that Repast is entirely built in Java, it supports a test-driven simulation development [11], i.e. unit tests can be carried out to validate the behaviour of a model. Moreover, NetLogo, Repast, and FLAME support empirical validation of models by collecting the generated time series output (or other data) from the simulation and then compared with collected data from the real world. Finally, the main unit of FLAME is an X-machine which is supported, it is possible verify and test the agent models individually.

3.4 Visualisation

An interesting concept about the visualization tool in NetLogo and Repast is the fact that they produce output in real time. FLAME on the other hand, requires that a model is initially run for a specific number of iterations, producing a textual output. This textual output can be later visualized. Comparing the forms of output, Repast can provide a histogram, visual display of the interactions between the agents, or a chart of time-series data. On the other hand, NetLogo

supports the following visualization options: line, bar, scatter plots, and visual display of the interactions between the agents as 2D and 3D models. FLAME behaves poor in this aspect, i.e. it supports only visual interactions display.

3.5 Supported Size of MAS and HPC

NetLogo supports hundreds or thousands of agents to operate independently, providing a clear picture of the micro-level behaviour of the agents, as well as the macro-level patterns (emergence) within the whole system [7]. NetLogo, however, does not support high performance computing. Repast Simphony is very similar to NetLogo in this aspect. There is another release of this framework, Repast for High Performance Computing (Repast HPC) intended for large-scale distributed computing platforms [2]. Finally, FLAME has a clear advantage by supporting simulations that contain up to millions of agents. Furthermore, only FLAME is designed to utilise high performance computing, and has been tested to work on both serial and parallel systems [6].

	NetLogo	Repast Simphony	FLAME
Modelling and Design	• Does not enforce model design.	 Does not enforce model design. Models can be built by utilizing point-and-click flowcharts. 	• Enforces designing of a model.
Implemen- tation	tions natively.Agent have embedded position and direction.Agents can be assigned to a position randomly.	 Supports movement functions natively. Agent have embedded position and direction. Agents can be assigned to a position randomly. External language interface: Groovy or Java. 	 ment functions natively. Agent do not have embedded position and direction. Agents have to be individually instantiated in a set
Verification Validation	Not supported. Empirical validation.	Not supported. Empirical validation.	Not supported. Empirical validation.
Testing Visualisation	 Not supported. In real time and directly executable as a visual simulation. Supports visual interactions display, line, bar, and scatter plots. Very good display. It is embedded. 	Supported. • In real time and directly executable as a visual simulation. • Supports histogram, visual interactions display and a chart of time-series data. • Very good display. • It is embedded.	 Not supported. Not in real time. Is not directly executable as a visual simulation. Supports only visual interactions display. Very poor display. Comes as an external tool.
Supported Size of MAS HPC	Hundreds or thousands of agents. No.	Hundreds or thousands of agents.	Millions of agents. Yes. Works on both serial
	110.	designed for HPC.	and parallel systems.

Table 2. Comparison between NetLogo, Repast and FLAME

4 Conclusion

This work provides a survey of visualization approaches most suitable for statebased MAS with spatial characteristics, i.e. NetLogo, Repast and FLAME. Based on the summary presented in Table 2, it might be concluded that NetLogo and Repast are very close to each other when it comes to modelling. FLAME has an advantage in this aspect, because it enforces the modeller to think in terms of design. Other advantages of FLAME are the support of HPC and the supported size of MAS, which extends to millions of agents. NetLogo and Repast have advantage over FLAME in terms of visualization and implementation. These platforms produce real time visualization and various display of the data, including charting and statistical data analysis. Moreover, spatial characteristics such as movement functions, or the position and direction of an agent, are natively supported.

Future work includes improving some of the disadvantages found in FLAME. This consists of support for spatial properties and a tool for automatic generation of set up data.

References

- 1. Robertson, D.A.: Agent-based modeling toolkits. Academy of Management Learning and Education 4(4) (2005) 525–527
- Collier, N.T., North, M.J.: Repast SC++: A platform for large-scale agent-based modeling, Large-Scale Computing Techniques for Complex System Simulations, Wiley (2011) (In Press).
- 3. Minar, N., Burkhart, R., Langton, C., Askenazi, M.: The swarm simulation system, a toolkit for building multi-agent simulations. Technical report (1996)
- 4. Wilensky, U.: Netlogo simulation software. Center for Connected Learning and Computer-Based Modeling (1999)
- 5. Luke, S., Balan, G.C., Panait, L., Cioffi-Revilla, C., Paus, S.: Mason: A java multiagent simulation library. Proceedings of the Agent 2003 Conference (2003)
- Pogson, M., Smallwood, R., Qwarnstrom, E., Holcombe, M.: Formal agent-based modelling of intracellular chemical interactions. Biosystems 85 (2006) 37–45
- 7. Wilensky, U.: Netlogo segregation model. Center for Connected Learning and Computer-Based Modeling (1997)
- 8. Holcombe, M.: X-machines as a basis for dynamic system specification. In: Software Engineering Journal. (1988) 69–76
- Bonabeau, E.: Agent-based modeling: methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences (2002) 7280–7287 Washington, United-States.
- Gurcan, O., Dikenelli, O., Bernon, C.: Towards a generic testing framework for agent-based simulation models. Federated Conference on Computer Science and Information Systems (FedCSIS) (Sept. 2011) 635 – 642 Emory University, Atlanta.
- 11. Collier, N.T., Howe, T.R., North, M.J.: Test-driven simulation development using repast simphony. Proceedings of the North American Association for Computational Social and Organizational Science (NAACSOS) (June 2007) Emory University, Atlanta.

A thermographic Atlas of the Human Body

Viktoria Rajtukova¹, Monika Michalikova¹, Teodor Toth¹, Jozef Zivcak¹

¹Technical University of Kosice, Faculty of Mechanical Engineering, Department of Biomedical Engineering and Measurement, Letná 9, 042 00 Kosice, Slovakia

{viktoria.rajtukova, jozef.zivcak, monika.michalikova, teodot.toth}@tuke.sk

Abstract. The presented study deals with the temperature distribution of healthy human bodies measured using infrared thermography. A database of 540 thermograms from different positions or localities of the human body was created. The following eighteen views of the human body were captured: TBA (total body anterior), TBD (total body dorsal), TBAA (total body anterior abduction), TBDA (total body dorsal abduction), TBR (total body lateral view right), TBL (total body lateral view left), CA (chest anterior), UB (upper back), RAA (right arm anterior), RAD (right arm dorsal), LAA (left arm anterior), LAD (left arm dorsal), RHA (right hand dorsal), RHP (right hand plantar), LHD (left hand dorsal), LHP (left hand plantar), FA (face) and FP (foot plantar).

Keywords: thermography, methodic, infrared thermograph, reference sample, atlas

1 Introduction

The goal of this paper is to create a database of normal thermograms of young healthy people. The database could in the future be helpful in the diagnosis of various diseases such as, for example, breast cancer, vascular diseases and skin diseases, and in studies of inflammatory responses, sleep research and pain related thermal dysfunctions [1],[2], [3], [4].

Physiological temperature distribution means temperature distribution which is measured from some position and locality on a healthy human [5],[6].

The technique uses an infrared imaging and measurement camera to "see and measure" invisible infrared energy being emitted from an object. It is a tool for the study of surface temperature distribution of living organisms [4].

The first publication on medical thermography, entitled "Standardization of thermography in diseases of musculoskeletal system - the recommended procedure,"

appeared in 1978. The study looked at the basic requirements for a methodology that was approved by the European Committee of Rheumatologists and Radiologists [7].

In 1988, the European Association of Thermography issued is "Evaluation of Thermography". The study captures the essential points of analysis and measurement conditions [7].

In 2010 "New standards for devices used for the measurement of human body temperature "were issued dealing with the development of thermometers and devices for measuring the temperature of the human body and a description of the new standards, which have now gained international recognition. The study also describes the practical application of these techniques in clinical medicine [8].

The atlas of normal thermograms is a database of measurements, taken from physiologically healthy human volunteers with the use of infrared thermography [4], [6].

The resulting images of each group (e.g. all arms of 20 to 30 year old obese males) are aligned in identical positions and magnifications using simple image processing techniques. They are then 'morphed' into a standard shape and superimposed onto each other so that statistical parameters such as the mean and standard deviation of each point of the body section in question can be calculated. The collection of 24 body sections is the "Infrared Atlas of Normal" [9].

2 Material and methods

2.1 Conditions of Measurement

The room in which the thermographic measurements are made has to meet certain conditions, including a constant room temperature and a certain level of humidity and, light, and must have the required room equipment.

In this case the examination room had air conditioning, which allowed us to achieve the necessary room temperature of 22.5 degrees Celsius ($\pm 1.7^{\circ}$ C). The room temperature is higher or lower, the examined person could end up sweating or shivering; thus disputing the measurements. The room temperature was controlled by TESTO 810 (a temperature measuring instrument with an infrared thermometer). Having window blinds is an advantage. If we perform these kinds of measurements in daylight, we get (except for a thermogram) a classic photograph. There can be no heater in or close to the room. Metal objects and reflective surfaces must also be removed.

The thermovision camera has to be sufficient distance from the measured person (approx 2 m), this means a larger room is needed to do this. The optimal room dimensions are 3x4 m [7].

2.2 Methodology of sensing system preparation

Fluke IR-Fusion technology captures a visible light image in addition to the infrared image. It joins the two images and makes a single image or makes a combination of a picture in a picture. IR-Fusion technology helps us to identify and record suspicious components and helps us to repair them properly the first time. The head of the camera with the lens rotates within 180 degrees and allows us to show and record images in areas that are hard to reach. SmartView software provides us with tools for analyzing infrared images, adding explanations and preparing reports. It allows us to edit these reports in a way that will satisfy specific workflows and their requirements.

Before measuring, the room temperature and emissivity have to be adjusted on the Fluke Ti 55-20 and the right kinds of lenses have to be chosen. The emissivity value for a human body is 0.98. The emissivity value is determined from a "Table of emissivity" freely available on the internet [6], [10].

2.3 Methodology of thermovision somathometry

With a modern thermovision camera there needs to be few second for it to start to work. First of all, the camera has to be set at an appropriate distance from the scanned subject. To have a thermogram on the entire screen of the thermovision camera, it is necessary to decide on an appropriate distance, height and angle of measurement. It should be ensured that while measuring a subject, the image should not be exposed to heaters, reflective components or any other such subjects.

We have to be able to see the visible end points on the screen of the thermovision camera.

Software packages distributed with the camera are used for processing the obtained thermograms. Software allows us to archive images made with thermovision camera.

2.4 Methodology of subject preparation

The day before the measuring the subject has to be instructed to:

- The body should be dressed in loose clothing. Tight clothing on the skin leaving body prints and causes a reduction in blood flow to the body
- Subject should not eat food or smoke 4 6 hours before measuring
- · Subject should not drink alcoholic drinks and not take drugs
- Subject should not drink warm drinks at least 2 hours before the measuring
- Subject should not take part in any physiotherapy treatment
- Subject should not practice anything that needs increased physical activity 6 hours before measuring (running, swimming)
- come without any make-up and not have any kind of lotion applied
- subject need to be in a state of mental and physical condition (sample of healthy subject for this study)

• subject is informed about the way the measuring is done, informed about painless non-invasive handling and the safety measurement method

It is important to complete the "Questionnaire" by which we find out needed the information necessary for the statistical analysis of the thermovisual measuring. Thermal stabilization of the subject consists of acclimatization which takes 15 - 20 minutes. Before the measuring the subject is informed about the positions that will be captured. For the "thermographic atlas of human body" we captured 18 body positions.

2.5 Reports

Each report includes:

- date and time of the thermal image
- information about the subject
- technical parameters of the camera.

The color scale can be set before the measuring. In this experiment we set it to the grey scale because we wanted to focus on the details of the human body.

By processing the range of temperatures we find isotherms which help identify temperatures for the selected polygon.

Using the Smart View software we can detect changes of temperature on the surface of the human body. We can detect changes in symmetries on different parts of the body and deviations from the standards. Based on the measurements and other personal information about the subject, we can provide a general description of a thermogram and make an evaluation of it together with a histogram.

On an evaluated thermogram we can see signs determining the center position, see the center frame and point to the warmest and the coldest points on the body. These signs can be turned over this means we can set the sign which we want to have or do not want to have on the thermogram.

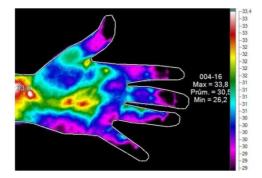


Fig. 1 Thermogram

3 Results and conclusions

Thermography enables us to measure the temperature of objects on the surface without any contact, to display it on a computer and to differentiate temperature fields.

Today, attention is paid to thermography worldwide because we want to start using thermography in everyday clinical practice. This project is following up on worldwide research. Its aim is to create a database of "normal thermograms" made with healthy people aged 16 to 28 living in Slovakia (mild temperature zone), with "normal anthropological parameters" like height, weight and body proportions.

Our database consists of 30 physiological healthy volunteers (n1=30; 15 male, Mn1=15; 15 female, Fn1=15). These subjects were captured by infrared camera in 18 different postures (1...18n30). We were able to obtain 540 thermograms (nTG=540). The average age of the volunteers was 25 years.

In Fig. 2 there is a comparison of average, maximum and minimum temperatures for each view displayed. The temperatures are calculated (arithmetic mean of all measured subjects in the selected projection) from marked body region, for example see Fig. 1.

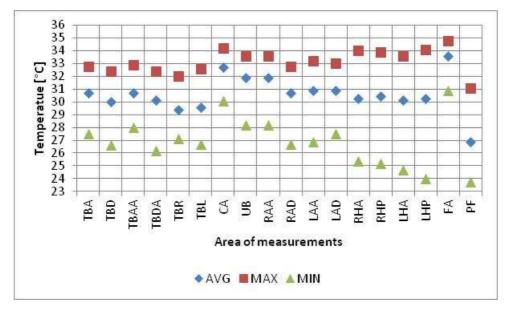


Fig. 2 Comparison of the temperatures.

The outcome of this study is a database of normal thermograms and a review of physiological temperature distribution.

The temperature values obtained can be used as a reference for medical thermography, where positive/negative findings of the disease or injury can be quantitatively assessed.

A thermographic Atlas of the Human Body

Table 1. Characteristics of the processing location TBA and TBD

Code	TBA - total body anterior view				
TBD - total body dorsal view					
	<i>f the image</i> : the most cranial point of the head <i>f the image</i> : soles of the feet				
palms (TBA) or	ding straight, standing on a pad, arms and legs slightly abducted, dorsum (TBD) of the hands point forwards, head is a vertical tated nor tilted to the side				
300 200 200 200 200 200 200 200 200 200	$\begin{array}{c} 027-1 \\ Max = 35,6 \\ Prům = 30,4 \\ Min = 21,9 \end{array} \\ \begin{array}{c} 33.6 \\ -34 \\ -34 \\ -33$				

Table 2. Characteristics of the processing location TBAA and TBDA

Code	TBAA - total body anterior abduction
	TBDA - total body dorsal abduction
11 0 0	<i>f the image</i> : the most cranial point of the head <i>f the image</i> : soles of the feet
•	ading straight, standing on a pad, legs slightly abducted, arm are ded, palm (TBAA) or dorsum (TBDA) outstretched and fingers

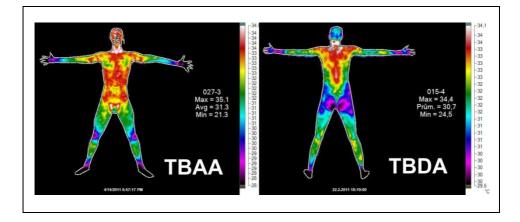


Table 3. Characteristics of the processing location TBR and TBL

	1				
Code	TBR - total body,	right lateral view			
TBL - total body, left lateral view					
11 0 0	f the image: the mos f the image: soles of	st cranial point of the head			
•		ng on a pad, legs are side by side. Arms scissor,			
1		camera, fingers extended, head is a vertical			
position, neither ro	tated nor tilted to the	le side.			
P	position, neither rotated nor tilted to the side.				

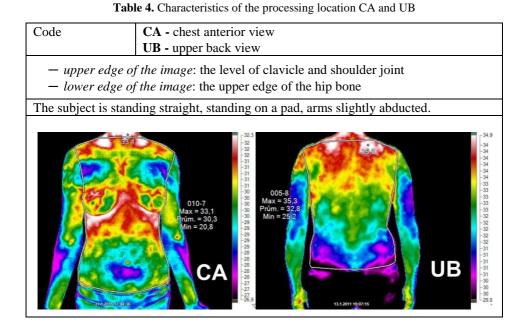


Table 5. Characteristics of the processing location FA

Code	FA - face	
cranial point — lower edge	e of the image: the most nt of the head of the image: below the level of hyoid bone	023-17 Max = 35.8 Prùm, = 34,1 Min = 25,0
position, head	terested in a seated is a vertical position, poking straight ahead.	

Table 6.	Characteristics o	f the processing loo	cation RAA, RAD,	LAA and LAD
----------	-------------------	----------------------	------------------	-------------

(Code	RAA - right arm anterior, LAA - left arm anterior, RAD - right arm dorsal, LAD - left arm dorsal,
	 upper edge of the image: tip of the middle finger (the end point of the longest fingers in the upper extremity) 	
	- lower edge of the image: below the axillary fold.	

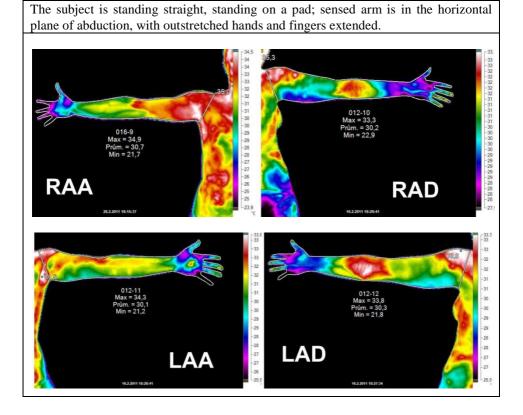


Table 7. Characteristics of the processing location RHD, RHP, LHD and LHP

Code	RHD - right hand dorsal, LHD - left hand dorsal, RHP - right hand plantar, LHP - left hand plantar,
 upper edge of the image: 1 – 2 cm over carpal joint lower edge of the image: tip of the middle finger (the end point of the longest fingers in the upper extremity). 	
The subject interested	in a seated position. Sensed hand is placed on a substrate, with

outstretched and fingers extended.

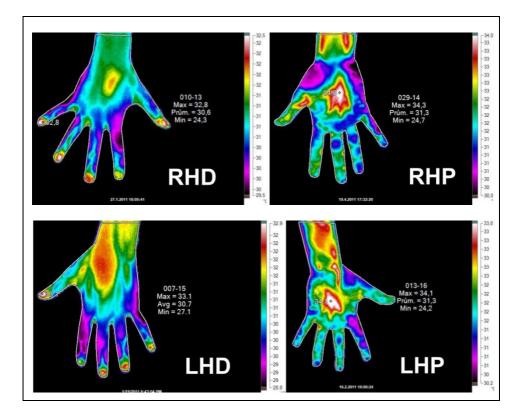


Table 8. Characteristics of the processing location PF

Code PF - plantar foot	
 <i>upper edge of the image</i>: the end point of the longest finger to the foot leg, tip of the great toe <i>lower edge of the image</i>: The final selection of heels 	001-18 Max = 34.0 Prim. = 30.5 Min = 25.0
The subject interested in a seated position. Lower limb landed on the mat, feet outstretched, slightly in abduction.	PF

Acknowledgment

This contribution is the result of the project implementation: Center for research of control of technical, environmental and human risks for permanent development of production and products in mechanical engineering (ITMS: 26220 120060) supported by the Research & Development Operational Programmers funded by the ERDF.

Reference

- Ammer, K.: Thermology 2009- a computer- assisted literature survey; In: Thermology International; Published by the Austrian Society of Thermology and European Association of Thermology; Vol.20, No.1; 2010; p.5-10; ISSN 1560-604X.
- [2] Jain, R.K., Gullino, P.M.: Thermal Characteristics of Tumors: Applications in Detection and Treatment, Annals of the New York Academy of Science, Volume 335, ANYAA 9 335 1-542, 1980
- [3] Kerr, J.: Review of the effectiveness of infrared thermal imaging (thermography) for population screening and diagnostic testing of breast cancer, NZHTA Tech Brief Series, July 2004 Volume 3 Number 3
- [4] Ring, E. F. J.: The historical development of temperature measurement in medicine, Infrared Physics & Technology, Volume 49, Issue 3, 2007, Pages 297-301
- [5] Ring, E. F., The historical development of thermometry and thermal imaging in medicine, Journal of Medical Engineering & Technology, 2006 Jul-Aug;30(4):192-8
- [6] Infrared Atlas of Normals, A Data Base of Normal Human Skin Temperature Distribution, http://www.comp.glam.ac.uk/pages/staff/pplassma/MedImaging/Projects/IR/ Atlas/index.html.
- [7] Hudák, R.: Termografická diagnostika v procese rehabilitácie paraplegických a tetraplegických pacientov. Doktorandská práca. Košice: Technická univerzita v Košiciach, Strojnícka fakulta, 2007. 144s.
- [8] Ring, E.F., McEvoy, H., Jung, A., Zuber, J., Machin, G.: New standards for devices used for the measurement of human body temperature. J Med Eng Technol. 2010 May; 34(4):249-53. Review. PubMed PMID: 20397848. http://www.ncbi.nlm.nih.gov/pubmed/20397848
- [9] Ring, E.F.: Quantitative thermal imaging, Clin. Phys. Physiol. Meas. 11 87-95, cit. 2010-12-12; http://www.medimaging.org
- [10] Spectral emissivity of skin and pericardium http://iopscience.iop.org/0031-9155/18/5/307

A three-dimensional cellular automaton for network traffic simulation

Stylianos Dimitriou*,

Technological Educational Institute of Serres, Serres, Greece, sdimitr@gmail.com

Abstract. Network simulators are software tools used by researchers in order to propose and evaluate new protocols. Developing a robust network simulator is an intricate task since it requires appropriate design models capable of reproducing the characteristics of a network. Cellular automata are computational models that can provide the functionality required to implement a network simulator. Although simulating network traffic with cellular automata is not an entirely new idea, current proposals have several limitations concerning the network topology that they can be applied, as well as the protocol stack. This paper presents a new network simulator application using a three-dimensional cellular automaton. The proposed model is not limited by the number of communicating nodes or by the probable interconnections among them; however protocol implementation and traffic pattern analysis is still at an early stage.

Keywords: Computer networks, network simulation, cellular automata.

1 Introduction

One of the many challenges computer network engineers face is proper evaluation of new protocols in a variety of conditions as close as possible to real-life networks. Indeed, integrating a new and unevaluated protocol in an existing network is usually avoided, since it is a costly and risky operation with unpredictable results. Protocol evaluation using a testbed infrastructure is also a costly procedure as, in order to emulate actual network conditions, several expensive communicating devices are required. Network simulators possess several advantages over hardware-oriented methods since they do not require costly equipment, can simulate a variety of topologies and conditions using only one computer, are relatively faster and allow easier monitoring. Some of the most widely used network simulators today are ns-2/ns-3 [5], OPNET [6] and NetSim.

Designing a new network simulator application is a complicated procedure; yet sometimes necessary in order to extend the functionality of existing solutions to new

^{*} Corresponding author: e-mail: sdimitr@gmail.com, address: Technological Educational Institute of Serres, end of Magnesia street, 62124, Serres, Greece.

network types and applications. Network simulators are usually built as event-driven discrete modular applications. Cellular automata (CA) are computational tools that can provide the functionality required to implement a network simulator. They are computational models that can simulate time-dependent phenomena as temporally and spatially discrete systems. CA are structured as grids of cells (usually two-dimensional) where each cell has a finite number of states and is updated in a discrete fashion based on specific rules. Since several processes can be modeled as discrete systems, CA have a wide range of applications, such as the Game of Life [2], fluid flow, sand pile simulation, earthquakes [1] and forest fire spread [3].

In this paper, a preliminary form of a network simulator using a cellular automaton is presented. The proposal departs from the original NaSch [4] traffic model and adapts it to computer networks. The novelty of this idea is the introduction of a threedimensional CA which can simulate almost any network topology, as long as the links between the nodes are clearly defined. Although the simulator is still under development, there is a core functionality that can be used for real-life applications.

The rest of the paper is organized as follows. Section 2, presents the simulator model and the implemented functionality. In section 3, a basic set of experiments is conducted so as to make an initial validation of the simulator. Last, section 4 concludes with remarks concerning further development of the application.

2 The network traffic simulator

A typical computer network consists of nodes and links (Fig. 1). A node is the communication device of the network and can be a packet sender, a packet receiver or a packet forwarder. A link is the wired or wireless connection between two nodes. Both links and nodes should be modeled in the CA.

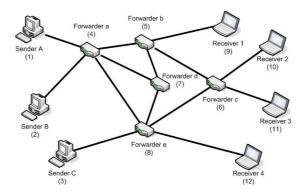


Fig. 1. A generic network topology with 12 nodes and 14 links.

Before delving into the CA model, current characteristics and limitations of the simulator will be mentioned.

• Links. Links are simplex by default. This means that data travels only to one direction. In order to use duplex links, the inverse link should be defined as well. Links have three parameters: propagation delay (ms), maximum packet

transmission rate (packets/ms) and packet loss rate (%). For simplicity, all propagation delays are assumed integer.

- Senders. Senders send packets in the network. At present, only Constant Packet Rate transmission is implemented. Packets have no specific receiver, they just flood the network. Senders have three parameters: packets sent per transmission (packets), interval between two transmissions (ms) and fail probability (%). Fail probability is the probability that the node will not function at all.
- **Receivers**. Receivers receive the packets forwarded to them by the network. No acknowledgements are sent. They have one parameter: fail probability (%).
- **Forwarders**. Forwarders forward the packets to the network. Since packets have no specific recipient, forwarders divide the packets equally to their outgoing links. They have two parameters: buffer size (packets) and fail probability (%).

One might argue about the purpose of a network simulator where packets have no specified receivers. Most network simulators are connection-oriented, meaning that they conduct packet-level monitoring and traffic control. However, there are several cases when we are not interested in specific users but on the network as a whole, such as when we want to detect the bottleneck links of a network or examine how adding and deleting links will affect the traffic. In these cases, focusing on the total traffic of the network, instead of individual flows, increases speed without losses in reliability.

Links are modeled by the original NaSch model for traffic simulation. This model consists a one-dimension CA and considers random accelerations and decelerations of the traffic. Since the conditions in the network links are simpler, the NaSch model will be simplified (Fig. 2). The update rule of a link will be a(i,t+1)=a(i-1, t). The internal state a(i,t) is less or equal to the maximum packet transmission rate and the length of the CA is equal to the link's propagation delay. The boundary conditions will be determined later.

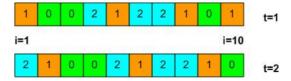


Fig. 2. The CA which models a network link.

Modeling a node is easier, since it requires only one cell. Thus, the entire network can be modeled as a three-dimensional CA (Fig. 3 (a)). If N is the total number of nodes and D is the maximum propagation delay of all the links, then the dimensions of CA are N*N*D.

If we consider the 3D structure of Fig. 3a with coordinates (i,j,k), then node M $(1 \le M \le N)$ corresponds to the cell with coordinates (M,M,1). The link which connects nodes M and K with direction from M to K, corresponds to the column with cells from (M,K,1) to (M,K,d) where d is the propagation delay of the link and $1 \le d \le D$. The rest of the cells are never used.

Since it is hard to visualize the CA for a complicated topology, a simple example of a two-node network is presented in Fig. 3b. The first node is a sender (blue cell, (1,1,1)) and the second a receiver (red cell, (2,2,1)). A simplex link connects these

two nodes (green cells, (1,2,1) to (1,2,5)). Packets are generated in cell (1,1,1) and enter the link via cell (1,2,1). They travel through the link and arrive to the receiver. If there are forwarders in the network with more than one incoming links, incoming packets from all links will be added and saved in buffer.

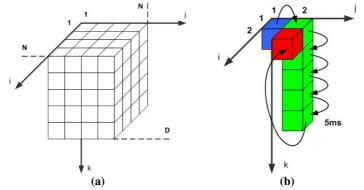


Fig. 3. (a) The structure of a generic network. (b) The CA for a two-node topology.

3 Experimental validation

In order to make an initial validation of the simulator, a small set of experiments has been conducted using a dumbbell topology with 6 nodes (Fig. 4). Each experiment ran for 100ms. The simulation was executed three times, with variable bottleneck link packet loss rate. The characteristics of the network can be seen below.

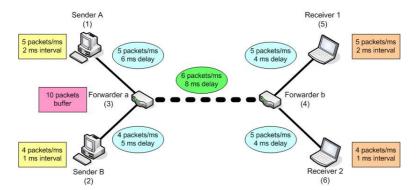


Fig. 4. A simple dumbbell topology with loss conditions for the bottleneck link.

The results of the three executions can be seen in Table 1. Even though the topology is elementary, some useful conclusions can be extracted. First, from examining where we have dropped packets due to overflow, we can detect the bottleneck links. If we know a priori the average expected sending rates of the senders, we can adjust the characteristics of the bottleneck link so that no congestion

occurs. Additionally, in case of wireless links we can see how loss conditions can affect the network traffic. In our case, 10% is loss is acceptable; however loss rate more than 20% has serious consequences in the smooth operation of the network.

Number of execution	First	Second	Third
Bottleneck link Packet Loss Rate	0%	10%	20%
Total packets sent	650	650	650
Total packets received	602	554	464
Total packets lost	48	96	186
Packets sent from node 1	250	250	250
Packets sent from node 2	400	400	400
Packets dropped from node 3 due to overflows	48	48	48
Packet lost at link 3-4 due to loss conditions	0	48	138
Packets received by node 5	352	309	244
Packets received by node 6	250	245	220

Table 1. Simulation results.

4 Conclusion and Future Work

In this paper, a new network simulator based on a three-dimensional cellular automaton has been proposed. Its features and limitations were presented and a small set of validating experiments were conducted. Although still under development, the simulator seems to perform properly under various conditions.

The most important extensions of the simulator will be the integration of more traffic patterns. Senders will be able to generate TCP-like traffic patterns and other UDP applications beyond constant packet rate should also be considered. Moreover, the addition of several packet sizes should be included, since bulk data traffic tends to utilize bigger packet sizes than real-time traffic. Thus, these dynamics should be considered. Of course, proper validation of the complete model of the network simulator is essential for real-life deployment where we are interested not only for speed, but also for reliability and consistency.

References

- 1. Castellaro, S., Mulargia, F.: Implementing cellular automata models for earthquakes on parallel computers. Geophysical Research Letters, vol. 30, 1204, 4 pp. (2003)
- Chopard, B., Droz, M.: Cellular Automata Modeling of Physical Systems. Cambridge University Press (1998)
- Hernández Encinas, L., Hoya White, S., Martín del Rey, A., Rodríguez Sánchez, G.: Modelling forest fire spread using hexagonal cellular automata. Applied Mathematical Modelling, vol. 31, iss. 6, pp. 1213-1227 (2007)
- 4. Nagel, K., Schrekenberg, M.: A cellular automaton model for freeway traffic. Journal de Physique I, vol. 2, no. 12, pp. 2221–2229 (1992)
- 5. Network Simulator, http://www.nsnam.org/
- 6. OPNET, http://www.opnet.com/

The Challenge of Engineering Multi-Layer Monitoring & Adaptation in Service-Based Applications

Konstantinos Bratanis¹, Dimitris Dranidis¹, and Anthony J.H. Simons²

¹ South-East European Research Centre (SEERC), International Faculty, The University of Sheffield, 24 Proxenou Koromila, 54622, Thessaloniki, Greece, {kobratanis,dranidis}@seerc.org http://www.seerc.org/soe/

² Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield S1 4DP, UK, a.j.simons@sheffield.ac.uk http://www.sheffield.ac.uk/

Abstract. The existing monitoring and adaptation (M&A) approaches for service-based applications (SBAs) are highly fragmented and they are not effective when they operate in isolation. A recently proposed research direction towards addressing more complex and more effective adaptations is the engineering of multi-layer monitoring and adaptation (multi-layer M&A). Multi-layer M&A is concerned with the integrated operation of different M&A approaches, each one addressing a specific aspect at some layer of an SBA. In this paper we discuss the topic of multi-layer M&A. We present the current state of the art in multi-layer M&A and the current limitations. We argue that the current state of the art is significantly limited due to the lack of a concrete framework for engineering flexible, extensible, dynamic and effective multi-layer M&A approaches. We propose to address this gap through the investigation of a framework for engineering the integration and the coordination of diverse monitoring and adaptation mechanisms. Finally, we set the concrete research goals and objectives of the proposed research project.

Keywords: service-oriented computing, service-based applications, multilayer monitoring & adaptation, architecture, integration, coordination

1 Introduction

Service-based applications (SBAs) are typically viewed as layered applications that are developed by integrating multiple heterogeneous software-based services [1]. The constituent services are usually developed, controlled and provided by different third-party organisations, not necessarily by the organisation that owns the SBA. An SBA usually incorporates a big number of services, often consumed over a network, that are used as the building blocks for perform the desired functionalities of the application. The services are used to instantiate service compositions or business processes that realise business workflows.

SBAs are typically complex applications and therefore harder to manage during the design-time and the run-time as well, due to a number of reasons:

- 1. Service-orientation fosters the implementation of highly dynamic systems, where services can be discovered, consumed, and removed at run-time;
- 2. Services are often developed independently from the execution context, that is, a service can be consumed in an application that was never envisaged;
- 3. Services can evolve without prior notification, that is, services exposing an expected behaviour through an agreed interface can have their behaviour or interface modified;
- 4. The quality of service is variable at run-time, that is, even if a service was performing well during design-time, it is not guaranteed that will perform the same at run-time;
- 5. Service-based systems operate in highly heterogeneous and dynamic execution context.

Given the above, one of the most critical challenges associated with the dynamics and the complexity of SBAs is the development of approaches for engineering self-adaptation in such applications. This challenge stems from the high probability that the requirements satisfied during design time will be violated at run-time, ultimately leading to faults or total failure of the SBA. As such, the ability to detect, predict, diagnose, recover, compensate and prevent such violations has been widely acknowledged by the SOC research community as a significant and challenging problem [1, 2].

A wide range of fragmented monitoring and adaptation (M&A) approaches have been proposed in the literature for partially addressing very specific problems from a very particular perspective at a single layer of an SBA. It was recently recognised that the high fragmentation of M&A approaches does not facilitate the development of effective solutions for engineering adaptable SBA [1, 3]. The isolated operation of M&A mechanisms, which ignore the layered nature of the SBA, can lead to problems, such as the incorrect detection of adaptation requirements, or the enactment of ineffective adaptation actions.

In order to address the aforementioned multifaceted problem, new research directions were proposed towards developing multi-layer (or cross-layer) monitoring and adaptation (multi-layer M&A) approaches for SBA [3]. Multi-layer M&A is concerned with the integration of monitoring and adaptation mechanisms targeting multiple aspects across the layers of an SBA.

Despite the fresh directions, the current state of the art in multi-layer M&A has demonstrated few multi-layer approaches, which have been based on the adhoc integration of pre-existing monitoring and adaptation mechanisms. Such efforts realised multi-layer M&A from a particular perspective and they addressed a very specific problem. We argue that the current state of the art in multi-layer M&A is significantly limited due to the hard-wired approach to integration of

mechanisms and their uncoordinated operation in the existing multi-layer M&A approaches.

In order to overcome the current limitations, we propose the investigation of an integration and coordination framework for engineering multi-layer M&A solutions that are flexible, extensible, dynamic, and effective. In the rest of the paper, we briefly present the current research efforts on multi-layer M&A and the current limitations. Finally, we present the goal and the objectives of the proposed framework.

2 State of the Art in Multi-Layer M&A

Multi-layer M&A has been acknowledged as a promising direction for implementing adaptable SBAs [4]. However, the recent research efforts towards addressing the multi-layer M&A of SBA are rather scattered. In this section we present the outcome of a more comprehensive literature review, which we have conducted in a previous work [5]. We have classified the small number of existing research efforts into three groups: (i) research efforts addressing diverse aspects of multilayer M&A, (ii) initial efforts presenting ad-hoc integrations of pre-existing M&A approaches and (iii) efforts targeting the architectural perspective of multi-layer M&A.

Concerning the first group of research efforts, the monitoring aspects of multilayer M&A have been investigated in [6]. Efforts in [7,8] present approaches targeting the problem of deciding adaptation strategies. Also, research efforts such as the ones presented in [9,10] are concerned with the role of servicelevel agreements (SLA) in multi-layer M&A. The aforementioned research efforts address diverse aspects of multi-layer M&A.

Regarding the second group of research efforts, efforts such as [11–13] are the first to realise the integration of diverse M&A approaches for implementing multi-layer M&A approaches. The aforementioned research efforts have managed to realise an ad-hoc integration of M&A approaches, that is, they integrate a specific monitoring and analysis mechanisms with specific decision and adaptation enactment mechanisms.

The third and last group of research efforts, such as [14–16] explicitly address the architectural perspective of the integration of diverse M&A mechanisms for multi-layer M&A. Such efforts are rather scarce.

2.1 Current Limitations

The previous section presented the three groups of the small number of recent existing research efforts cantered around the topic of multi-layer M&A. The small amount of research efforts is a sign of the immaturity of the field, which should set as a top priority the development of methods for engineering the integration of diverse M&A mechanisms, as well as, their integrated coordination. We argue that the current state of the art in multi-layer M&A is significantly limited due to

the lack of such a framework, which addresses the integration and coordination of diverse M&A approaches.

Our argument is supported by the fact that all three research efforts [11–13], which constitute the current state of the art in multi-layer M&A approaches, appear to follow an ad-hoc integration of pre-existing monitoring and adaptation mechanisms. The hard-wired approach to integration employed by these efforts limits their capabilities in terms of extensibility, dynamism, and effectiveness. These efforts address multi-layer M&A from a very specific perspective, while they focus more on the operation of the individual mechanisms, without considering the effectiveness of the integrated mechanisms as a whole and across the layers of an SBA. Additionally, they leave aside the extensibility aspect concerned with the integration of other adaptation and monitoring mechanisms, which can be used for addressing other concerns of the cross-layer setting.

Moreover, the research efforts in [14–16], which constitute the current state of the art concerning the architectural aspects of multi-layer M&A, outline initial ideas towards addressing the challenges associated with the architectural aspects of multi-layer M&A. However, these efforts have only managed to scratch the surface of the problem. They only presented insights and partial solutions, which address a very narrow aspect of the architecture required for the integration and the coordination of diverse monitoring and adaptation mechanisms. Finally, these efforts ignore completely the issue of dynamism in multi-layer M&A, i.e. the ability to deploy or remove M&A mechanisms during run-time.

2.2 The Challenge of Engineering Multi-Layer M&A

In this paper, we put forward the challenge of engineering multi-layer M&A approaches that are flexible, extensible, dynamic and effective. The flexibility of such approaches consists in their support for extending their capabilities through the integration of mechanisms for monitoring, analysis, decision, and adaptation. Also, the dynamism in such approaches refers to the ability to deploy and remove M&A mechanisms at run-time. The effectiveness of such dynamic solutions relies on an additional control loop for coordinating the parallel operation of the integrated mechanisms.

The abovementioned challenge comprises three main problems that call for a concrete solution. The first problem is how to architect the integration of diverse mechanisms and ensure the dissemination of information among them. The second problem is what information needs to be conveyed among the mechanisms and how this information is standardised. The last problem is how to coordinate the operation of the integrated mechanisms, such that they operate in an effective way. Addressing all three problems would provide the baseline for engineering flexible, extensible, dynamic and effective multi-layer M&A approaches for service-based applications.

3 A Framework for Engineering Multi-Layer M&A

In order to advance the current state of the art in multi-layer M&A, this report puts forward the investigation of a framework for engineering the integration and the coordination of M&A mechanisms. The purpose of the engineering framework is to facilitate the development of flexible, extensible, dynamic and effective multi-layer M&A approaches for SBAs.

The proposed framework comprises three main pillars. The first pillar of the framework is concerned with the architectural aspects of the integration, i.e. how to enable the loosely coupled integration of the interacting monitoring and adaptation mechanisms, while at the same time ensuring the propagation of information between them. The second pillar is concerned with the conveyance of information among mechanisms, i.e. what kind of information needs to be conveyed among mechanisms, such that their coordination is possible. The last pillar is concerned with the coordination of the monitoring and the adaptation mechanisms, i.e. how to achieve near real-time operational behaviour in multilayer M&A that will be capable of detecting situations occurred in monitoring and adaptation mechanisms, such that coordination actions are performed. A more detailed description of each pillar is available in [5].

3.1 Research Goal and Objectives

The proposed research project targets the challenge of multi-layer M&A of service-based applications, more particularly the need for engineering the integration and the coordination of diverse M&A mechanisms, and proposes to address this challenge through a framework for engineering the integration and the coordination of such mechanisms. The framework will be based upon a flexible architecture for the loosely coupled integration of mechanisms; an expressive event schema for facilitating the conveyance of information between the mechanisms; and a coordination method for controlling the operation of the integrated mechanisms.

More precisely, the proposed research project must fulfil five objectives to achieve its goal:

- 1. To develop an architecture that is extensible enough to facilitate the integration of diverse M&A mechanisms and enable the propagation of information between these mechanisms. Such an architecture aims to accommodate the complexity of an environment that comprises multiple, possible heterogeneous, M&A mechanisms, which focus on a great variety of activities, from different perspectives and with varying concerns. The architecture should account for the dynamicity through the appropriate provisions for enabling the deployment and removal of mechanisms during run-time.
- 2. To develop an event schema that is expressive enough to facilitate the conveyance of information between M&A mechanisms and support the conveyance of feedback and coordination information. The event schema will

explicitly specify the events, which are expected to be produced and consumed by the integrated mechanisms. Also, it should support its extension through the inclusion of new events when new mechanisms are being integrated.

- 3. To develop a method for coordinating the operation of the monitoring and the adaptation mechanisms, in order to avoid undesirable behaviour of the integrated mechanisms as a whole. The coordination process should comprise the detection of situations calling for coordination, reasoning about the action required for coordinating the operation of the mechanisms involved in the situation, and disseminate the coordination information to the involved mechanisms.
- 4. To implement a prototype of the engineering framework in an open-source service-oriented middleware, in order to demonstrate a proof of concept implementation. The prototype should realise all three parts of the framework, i.e. the sketched architecture, the event schema, and the coordination method.
- 5. To evaluate the prototype implementation of the framework, demonstrating a proof of concept for supporting the integration and the coordination of diverse M&A mechanisms. The evaluation will be carried mostly through experimental simulation, while also the integration of concrete mechanisms will be explored, based on their availability.

4 Conclusion

In this paper we presented the few recent research efforts on multi-layer M&A of SBAs. We argued that these efforts are rather scattered. We explained the current limitations in the existing literature. In order to advance the current state of the art, we proposed the investigation of a framework for engineering multi-layer M&A approaches that are flexible, extensible, dynamic and effective. As future work we will develop the engineering framework based on the objectives that we have mentioned in this paper.

Acknowledgements We cordially thank the anonymous reviewers for their constructive and detailed comments on improving the paper.

References

- 1. Papazoglou, M., Pohl, K., Parkin, M., Metzger, A. eds: Service research challenges and solutions for the future internet: S-cube - towards engineering, managing and adapting service-based systems. Springer-Verlag, Berlin, Heidelberg (2010).
- Di Nitto, E., Ghezzi, C., Metzger, A., Papazoglou, M., Pohl, K.: A journey to highly dynamic, self-adaptive service-based applications. Automated Software Engineering. 15, 313341 (2008).
- Kazhamiakin, R., Pistore, M., Zengin, A.: Cross-layer adaptation and monitoring of service-based applications. Proceedings of the 2009 international conference on Service-oriented computing. pp. 325334. Springer-Verlag, Berlin, Heidelberg (2009).

- Marconi, A., Bucchiarone, A., Bratanis, K., Brogi, A., Ca'mara, J., Dranidis, D., Giese, H., Kazhamiakin, R., Lemos, R., Marquezan, C., Metzger, A.: Research Challenges on Multi-layer and Mixed-Initiative Monitoring and Adaptation for Service-Based Systems. Proceedings of the 34th International Conference on Software Engineering (2012).
- 5. Bratanis, K.: Towards an integration and coordination framework for engineering multi-layer monitoring and adaptation of service-based applications. Technical Report, South East European Research Centre (SEERC), Thessaloniki, Greece (2012).
- Mos, A., Pedrinaci, C., Rey, G.A., Gomez, J.M., Liu, D., Vaudaux-Ruth, G., Quaireau, S.: Multi-level monitoring and analysis of web-scale service based applications. Proceedings of the 2009 international conference on Service-oriented computing. pp. 269282 (2009).
- Popescu, R., Staikopoulos, A., Liu, P., Brogi, A., Clarke, S.: Taxonomy-Driven Adaptation of Multi-layer Applications Using Templates. Self-Adaptive and Self-Organizing Systems, IEEE International Conference on. pp. 213222. IEEE Computer Society, Los Alamitos, CA, USA (2010).
- Zengin, A., Marconi, A., Pistore, M.: CLAM: cross-layer adaptation manager for service-based applications. Proceedings of the 1st International Workshop on Quality Assurance for Service-Based Applications. p. 21. ACM Press (2011).
- Fugini, M., Siadat, H.: SLA Contract for Cross-Layer Monitoring and Adaptation. In: Rinderle-Ma, S., Sadiq, S., and Leymann, F. (eds.) Business Process Management Workshops. pp. 412423. Springer Berlin Heidelberg, Berlin, Heidelberg (2010).
- Bratanis, K., Dranidis, D., Simons, A.J.H.: SLAs for cross-layer adaptation and monitoring of service-based applications: A case study. Proceedings of the International Workshop on Quality Assurance for Service-Based Applications. p. 28. ACM Press (2011).
- Schmieders, E., Micsik, A., Oriol, M., Mahbub, K., Kazhamiakin, R.: Combining SLA Prediction and Cross Layer Adaptation for Preventing SLA Violations. 2nd Workshop on Software Services: Cloud Computing and Applications based on Software Services., Timisoara, Romania (2011).
- Guinea, S., Kecskemeti, G., Marconi, A., Wetzstein, B.: Multi-layered Monitoring and Adaptation. In: Kappel, G., Maamar, Z., and Motahari-Nezhad, H.R. (eds.) Service-Oriented Computing. pp. 359373. Springer Berlin Heidelberg, Berlin, Heidelberg (2011).
- Zeginis, C., Konsolaki, K., Kritikos, K., Plexousakis, D.: ECMAF: An Event-Based Cross-Layer Service Monitoring and Adaptation Framework. 5th Workshop on Non-Functional Properties and SLA Management in Service-Oriented Computing (NFPSLAM-SOC11) co-located with ICSOC 2011. Springer (2011).
- Gjørven, E., Rouvoy, R., Eliassen, F.: Cross-layer self-adaptation of service-oriented architectures. Proceedings of the 3rd workshop on Middleware for service oriented computing. pp. 3742. ACM, New York, NY, USA (2008).
- Bratanis, K., Dranidis, D., Simons, A.J.H.: An Extensible Architecture for Runtime Monitoring of Conversational Web Services. Proceedings of the 3rd International Workshop on Monitoring, Adaptation and Beyond / ECOWS 2010. pp. 916. ACM (2010).
- Popescu, R., Staikopoulos, A., Clarke, S.: An extensible monitoring and adaptation framework. In: Dan, A., Gittler, F., and Toumani, F. (eds.) Service-Oriented Computing. ICSOC/ServiceWave 2009 Workshops. pp. 314324. Springer Berlin Heidelberg, Berlin, Heidelberg (2010).

Open Data Portal based on Semantic Web Technologies

Martin Kostovski¹, Milos Jovanovik¹ and Dimitar Trajanov¹

¹ Faculty of Computer Science and Engineering, Rugjer Boshkovik 16, P.O. Box 393, 1000 Skopje, Republic of Macedonia martin kostovski@yahoo.com, {milos.jovanovik, dimitar.trajanov}@finki.ukim.mk

Abstract. Recent activities of governments around the world regarding the publication of open government data on the Web, re-introduced the Open Data concept. The concept of Open Data represents the idea that certain data should be freely available to the public, i.e. the citizens, for use, reuse, republishing and redistributing, with little or no restrictions. The goal is to make non-personal data open, so that it can be used for building useful applications which leverage their value, allow insight, provide access to government services and support transparency. These data can contribute to the overall development of the society, by both boosting the ICT business sector and allowing citizens a deeper insight into the work of their government. This recent rise in interest for Open Data introduced the necessity for efficient mechanisms which enable easy publishing, management and consumption of such data. Therefore, we developed an Open Data Portal, with the use of the technologies of the Semantic Web. It allows users to publish, manage and consume data in machine-readable formats, interlink their data with data published elsewhere on the Web, publish applications build on top of the data, and interact with other users.

Keywords: Open Data, Open Government Data, Semantic Web Technologies.

1 Introduction

In the last years, many governments around the world started initiatives for opening up government data which they collect and produce during their everyday activities, and making it available for their citizens to use, analyze and redistribute [1,2,3]. This data is provided in open, machine-readable formats, so it can be used from applications built on top of it, leveraging its value and providing the end-users, i.e. the citizens, with useful information, insight, analysis, e-services, etc. In order to expose this data on the Web and make it available to the end-users, the governments develop and use systems known as open data portals. With the data published on these portals, the citizens can explore the different aspects of the government, as well as contribute in the public decision making process [4,5,6,7]. By publishing government data on open web portals, the governments are giving it back to the citizens, which indirectly paid for their creation with their taxes in the first place. This data creates additional business value within the economy: both individuals and the ICT business sector can create innovative applications on top of the open government data, and provide services which create new business value and profit [2]. These are the main motivation behind recent research in the fields of Open Data and Linked Data.

Another research field which has been active in the last ten years is the endeavor called the Semantic Web. It represents an upgrade of the current Web, which represents a network of hard-linked documents (web pages with hyperlinks between them), towards a network of information linked up in a way which allows them to be easily processed by machines, on a global scale. This is an efficient way of representing data on the Web, as a globally linked database [8,9]. The Semantic Web aims toward enabling machines to understand the semantics, i.e. the meaning of information on the Web. This will allow the machines to perform more intelligent actions over the vast number of information and data available on the Web today [10]. The technologies of the Semantic Web provide mechanisms for achieving this: they provide a layered structure of frameworks and languages for structuring data into knowledge, by the use of domain specific ontologies [11].

The technologies of the Semantic Web can be used as an integrator across different information, contents and applications for giving more useful data and knowledge, as well as automating many different processes [12,13,14,15,16]. With this, their scope shifts from the Web into many other fields of ICT.

In this paper, we present our open data solution, called the Open Data Portal, which allows users to publish, manage and consume data in machine-readable formats, interlink their data with data published elsewhere on the Web, publish applications build on top of the data, and interact with other users. The application is built with the use of the technologies of the Semantic Web, exploiting their benefits. The Open Data Portal enables the users – both individual programmers and ICT companies – to create and publish their own innovative applications and services based on the data, leveraging the value of the data itself and creating business value.

2 Related Work

2.1 Open Data

Public institutions collect and produce vast amounts of data as part of their daily functioning within the government. This data has a tremendous social value in terms of enabling the citizens to participate in public decision making, to know and have access to government services, etc. Using this open government data, they can contribute to the functioning of government and the democratic structures of society as a whole. A good example would be the mutual communication and trust between the government and society developed through data supply by the government and application building by the society. These types of applications are a good way for supplying the government back with analysis and statistics which they can use to improve their processes and functioning [2]. There is also a huge economic potential within the data, that can be used by businesses which want to add more value to the raw data and thereby take greater market share [17,18,19,20].

As we mentioned before, a great number of countries from around the world started publishing open government data in the last few years. These countries include USA, UK, France, Spain, Germany, Holland, Denmark, Norway, Finland, Italy, Greece and Estonia, as well as Mexico, Canada, Australia, New Zealand and India. Many of them have their own official national data catalogs such as the UK [21], the US [22], Denmark [23], Finland [24], and Spain [25]. Besides these national open data initiatives, there are local initiatives such as those in London [26], Zaragoza [27], Piedmont [28] and Helsinki [29]. The World Bank also has an open data program [30].

Below we present a detailed description and analysis of the Open Data initiatives of the United Kingdom, the United States of America and the World Bank, as one of the most successful open data stories.

UK datasets. The United Kingdom has achieved a significant success in developing its open data portal, http://data.gov.uk/. It features with over 5.600 datasets from all central government organizations and a number of other public bodies and local authorities, which are published in various machine-readable formats, such as XLS, CSV, RDF and XML. The portal also allows downloading the metadata from the appropriate dataset in JSON format. Data can be filtered on several criteria, such as publishers or institutions, nation (England, Wales and Scotland), tags, types and it can also be sorted by relevance, title, ranking, etc. Some particularly interesting datasets available on the UK open data portal include: data about the government ICT projects; trade of goods in the UK analyzed in terms of industry, cost indexes, total number of apartments; data for capital income and expenditure of the Department for Communities and Local Government; consumption in the Intellectual Property Office; data for national public transport; ICT activities of businesses in the UK; weekly fuel prices; index of average earnings; statistics on the labor market; price indices for construction and costs; business investment; monthly statistics on energy; cost data from the Commission on Audit; index services; balance of payments; quarterly energy prices; financial statistics, etc.

US datasets. The US open data portal, http://www.data.gov/, is a leading portal that shows the way for democratization of data and introduces innovation. The data available on the portal is gathered from several locations within the government. The portal has over 305.800 datasets which are published in various machine-readable formats, such as XML, CSV/TXT and RDF. The datasets can be queried by keywords, by categories such as finance, insurance, education, or by public institutions. The datasets include: review of deposits; investment of companies and class information; information about government projects; federal data on business opportunities; foreign labor statistics; index of producer prices; annual data for electricity; business employment dynamics; projections of employees; industry program for funding; resources needed for the budget; consumer and business lending program for investment initiative, productivity and costs; projections for population, etc.

World Bank datasets. The World Bank has an open data portal as well. It has two APIs for providing access to different datasets: one for indicators – time series data – and one for projects – data in relation to the operations of the World Bank. Both APIs implement RESTful interfaces which enable users to perform queries over the available data, using selection parameters. This technology allows for use of the services from both applications and web browsers. The indicators and projects from the datasets are available in XML, JSON and ATOM formats. The indicator API provides access to more than 3.000 indicators, some even spanning 50 years in the past. The API for projects provides access to all of the World Bank projects, including both active and closed projects. Some of the most notable datasets include: World Development Indicators; indicators about business and enterprises; development goals; global and national education statistics; global and national gender statistics; statistics on health and nutrition, etc.

2.2 Semantic Web and Semantic Web Tools

In recent years, there has been a great progress in the development of tools for representing and consuming knowledge on the Web. These software tools work with the technologies of the Semantic Web and are targeting the processes of design, implementation and maintenance of various semantics and concepts of functioning. Below we describe some of the more advanced semantic web tools based on different technologies, which can be used for developing scalable, reliable and advanced semantic web applications and solutions.

Semantic Web Tools in C# and .NET. DotNetRdf is a semantic web tool based on C# and .NET, falling within the categories of APIs and programming environments [31]. The API supports RDF [32], RDFS [33] and SPARQL as semantic web technologies.

Virtuoso is a semantic web tool used as an RDF repository [34]. Besides the many programming languages available in Virtuoso, such as C, C++, Python, PHP, Java, JavaScript, Action Script, Perl, Obj-C, it also uses C#. It can be defined as a triple store, an RDF generator and a SPARQL endpoint. Virtuoso works with RDF, RDFS, SPARQL, OWL [35], GRDDL [36] and RDFa [37] as semantic web technologies.

RDF Validator is a semantic web tool used as a validator and a visualizing tool [38].

ReDeFer is used as a convertor [39], which can also be classified as a visual RDF generator tool. It works with RDF, RDFa and OWL as semantic web technologies.

RelFinder is used for visualization [40]. It is a tool which extracts and visualizes relationships between objects represented in RDF.

Semantic Web Tools in Java. Jena is a semantic web Java tool, which falls into the categories of triple stores and programming environments [41]. Jena works with RDF, RDFS, SPARQL, OWL and GRDDL.

Protégé is used as an editor and a development environment [42]. It is a semantic web tool that can be classified as a development environment, an editor, a validator and a visual tool which works with RDF, RDFS and OWL.

Another tool that is used in conjunction with Jena is the semantic tool JRDF [43]. It is a tool that is based on RDF and falls into the category of APIs.

Another tool which can be combined with Jena is Virtuoso, mentioned above. The RDF Validator and ReDeFer tools can be used with Jena, as well.

3 Solution Description

We built a platform for open data which aims towards enabling users to easily publish, manage and consume open data in raw, machine-readable formats, interlink the data with other data previously published on the Web, publish applications build on top of the data, and interact with other users. With this, the platform provides the mechanisms which enable innovation in both application and use of open data, by either independent developers or ICT companies. As we pointed out earlier, the software applications built over this open data provide multiple benefits for the society: they provide the citizens with more insight into the work of their government, they create new e-services for the citizens based on preexisting data, they adds new business value into the economy, etc. In this section, we will present the details of our solution.

3.1 Platform Modules

The platform has a modular architecture and consists of six modules:

- module for displaying the user datasets and publications intended for public use;
- 2. module for publishing and presentation of applications built by the users, which use the data available on the portal;
- 3. module for sharing ideas;
- 4. module for forum discussion;
- 5. module for blogging;
- 6. module for displaying the published linked data and the semantically annotated (FOAF) profiles of the users and their relationships;

The first module (Fig. 1) displays the user datasets and publications intended for public use. The common use-case scenario within this module is as follows: the users log on to the system and are given the option to publish their data on the portal. Then they enter additional information in order to meet the input criteria required by the application, and upload one or more files which contain the data which they want to publish. If the files are in an RDF format, then they are submitted for review by the administrators of the portal. If the administrators decide that this data can be published on the portal, they try to find relations between the data and the global linked dataset, and thereby form a knowledgebase. The newly found (if any) relations and connections of the files are saved in a repository in which the knowledgebase is located, and the files are published on the portal as linked data. This data is then available for other users and visitors of the portal to obtain and reuse.

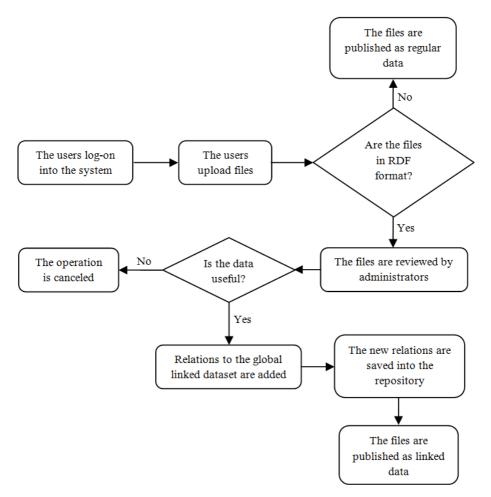


Figure 1: Use-case scenario of Module 1.

The second module (Fig. 2) is the module for publishing and presentation of applications built by the users, which use the open data published on the portal. The common use-case scenario for this module is as follows: first, the logged in users choose to publish a new application. Then they provide additional information to meet the input criteria for publication of the application, and enter the hyperlink pointing to the application, as well as a hyperlink pointing to the open data it uses. After the publication, the application becomes available for other users to browse and use.

The third, fourth and fifth module are places where the users can share ideas, discuss, and publish articles and other content on their blogs within the portal.

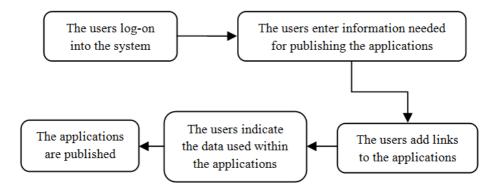


Figure 2: Use-case scenario of Module 2.

The sixth module is the module for displaying the published linked open data, and viewing the profiles of users in a semantically annotated format – FOAF [44]. This module provides a SPARQL endpoint [45] from which users can extract any data from the repository, i.e. the knowledgebase of the portal, by using SPARQL queries. The users can then convert the data into a desired format, such as DOC, XLS, TXT, XML, JSON or CSV. There is also the possibility of visualizing the data from the repository and finding semantic relations between different entities. The common use-case scenario for generating a FOAF profile and creating relations to other people is as follows: the logged in users choose to generate FOAF profiles and connect to other registered users. Then they enter additional information for filling the profiles and enter the links to FOAF profiles to people that they know. After filling in the information, the users generate FOAF profiles and their profiles are published on the portal (Fig. 3).

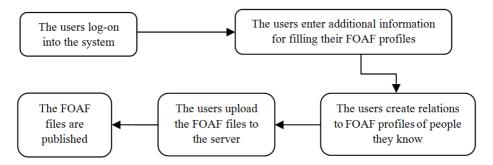


Figure 3: The use-case scenario for generating FOAF profiles of users.

3.2 System Architecture

Our solution consists of a presentation layer, an application layer, a business layer and a data layer, as shown on Fig. 4. The data layer was developed using Microsoft SQL Server, and the other layers were developed using Microsoft ASP.NET technology. The system

contains other elements as well, which are connected to the cloud: a semantic annotator placed within the business layer, an RDF repository represented by Virtuoso and placed within the data layer, and the DotNetRdf API which is placed within the application layer.

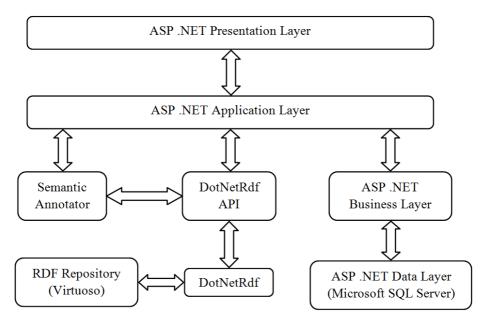


Figure 4: System architecture.

Semantic annotator. The annotator is used to upload RDF files on the server. The content of the file is semantically annotated in the appropriate format and the semantic data is registered in the repository which represents the knowledgebase of the system.

DotNetRdf API. The API provides data access to the repository by using SPARQL queries and data extraction. It extracts and stores RDF triples from the file that is uploaded to the Virtuoso repository. According to the base URI, DotNetRdf sets relations to the global linked dataset, stores them to the Virtuoso repository and updates the knowledgebase based on the relations between the entities.

RDF Repository. The RDF data is stored in the Virtuoso RDF repository as RDF Quads. One RDF Quad consists of a graph IRI, a subject, a predicate and an object within that graph IRI. According to that, RDF triples are taken from Virtuoso and relations are set to the global linked dataset. A knowledgebase is generated from these relations between the entities, and these relations are set in the Virtuoso repository.

Flow of data in the Open Data Portal. The process of data flow within the Open Data Portal for publishing RDF files and generating the knowledgebase is as follows:

- An RDF file is uploaded on the server and the file path is stored within a Microsoft SQL Server database;
- After the file is uploaded, RDF triples from the file on the server are stored in the Virtuoso repository, by using DotNetRDF. The base URI from the RDF file is saved, and according to it, RDF triples will later be selected from Virtuoso;
- After this, the RDF triples are selected from the Virtuoso repository with DotNetRDF and according to the base URI, and relations are set to the global linked dataset.
- If the relations are useful, they are stored in the Virtuoso repository with DotNetRDF, and the knowledgebase is updated, based on the relations between the entities.

The flow of data in the Open Data Portal for publishing RDF files and generating and updating the knowledgebase is shown in Fig. 5.

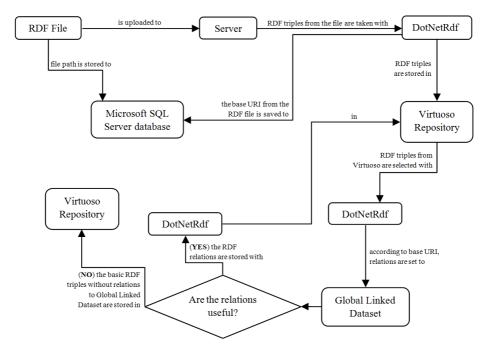


Figure 5: Flow of information in the Open Data Portal.

3.3 Advantages

One of the main motives for the development of the Open Data Portal is enabling the innovative users and businesses to leverage the value of open data, which is provided

by other users, companies or governments. The development of software applications and services over open data provides numerous benefits for the citizens and the society in general, such as greater transparency, communication, acquisition of trust, innovation, and business value.

The basic advantages our platform provides are:

- transparency, innovation and collaboration with citizens and end-users by sharing data, applications and ideas;
- ability for users to highlight and implement their ideas with the help of the open data provided by the Open Data Portal;
- ability to use open data within the business and supply various services to the end-users by developing applications based on the data, and generating new business value;
- ability to interact with other users and join the discussion by blogging and discussing on the forum;
- ability for users to connect with each other by exposing their accounts as FOAF profiles and making connections between them;

The Open Data principle used on this platform brings huge benefits for the end-users and citizens. This system provides a great social for the citizens and the governments, as well as economic potential which can be used by the ICT business sector to build and promote services and applications based on the open data from the Portal. With this, the end-users participate directly or indirectly in the use and consumption of the data, and can use the experience to gather better insight and understanding of the ways their government works. This experience can then be used as a feedback back to the government bodies, in order to provide better services for the citizens. This solution also enables and promotes transparency, innovation and openness to all users and provides assistance in the provision of better solutions and ideas.

3.4 Usage and Performance

The Open Data Portal was tested with 200 registered users. The performance was tested with the YSlow tool [46]. The publishing of data and applications is performed without delay to other processes and the time of publication is a few milliseconds. Finding relations to the global linked dataset does not have major impact on other processes in the system, as well. We tested this process by publishing RDF files with the size of 8MB, finding relations to the global linked dataset and forming/updating the knowledgebase. The problem of finding relations and its effect on system performance is solved through asynchronous communication and execution in the background, in order to reduce the impact on other processes. This is an acceptable solution, given the nature of the problem; we do not need the interlinked connection between the user data and the global linked dataset immediately. Therefore, no performance issues have been detected within the solution.

4 Conclusion and Future Work

The system and the application we designed and developed are intended to demonstrate the advantages the semantic web technologies provide when building an Open Data Portal. This platform should not be viewed as a platform which only offers an end product, but as a place which will constantly be upgraded and hopefully become a base which will offer opportunities for data and application creation, management, sharing and consumption.

Our initiative at the beginning is an attempt to encourage the government to open its doors by opening its data to the public, in order to provide more transparency, participation and collaboration. What we are trying to achieve is to build a strong community by removing the artificial and unfortunate separation of *swimming information*, caused by the fact that different government organizations keep their operating details and data on different online and offline locations.

One of the main objectives in further expansion is the development of applications for mobile devices based on the published open data. With these mobile applications, users will have greater access to services through their mobile devices and will be able to further utilize the benefits of open data.

Our future work will be focused on publication of useful open data, and encouragement of businesses and individuals to develop services and applications which will leverage the value of the published data. One such move would be the development of a module where users and companies will be able not only to publish, but to sell their applications and services. Thus, users and companies will be motivated to produce a great variety of innovative applications and help in the further development of the idea behind the Open Data Portal.

References

- 1. Berners-Lee, T., Shadbolt, N.: There's gold to be mined from all our data. The Times (2012)
- 2. Kundra, V.: Digital Fuel of the 21st Century: Innovation through Open Data and the Network Effect. Joan Shorenstein Center on the Press, Politics and Public Policy, Harvard College (2012)
- 3. Open Government Partnership, http://www.opengovpartnership.org/
- 4. Ding, L., Difranzo, D., Graves, A., Michaelis, J.R., Li, X., McGuinness, D.L., Hendler, J.: Data-Gov Wiki: Towards Linking Government Data. In: AAAI Spring Symposium: Linked Data Meets Artificial Intelligence (2010)
- Ding, L., Difranzo, D., Graves, A., Michaelis, J.R., Li, X., McGuinness, D.L., Hendler, J.: TWC Data-Gov Corpus: Incrementally Generating Linked Government Data from Data.gov. In: 19th international conference on World Wide Web (2010)
- 6. Sheridan, J., Tennison, J.: Linking UK government data. In: WWW Workshop on Linked Data on the Web (2010)
- 7. Ding, L.; DiFranzo, D.; Magidson, S.; McGuinness, D. L.; and Hendler, J.: The data-gov wiki: a semantic web portal for linked government data. In: International Semantic Web Conference (2009)
- 8. Palmer, S.B.: 2001-09. The Semantic Web: An Introduction. http://infomesh.net/2001/swintro/

- 9. Yu, L.: Introduction to the Semantic Web and Semantic Web Services. Taylor & Francis Group, LLC (2007)
- 10. What is Semantic Web, http://www.altova.com/semantic_web.html
- Antoniou, G., Harmelen, F.V.: A Semantic Web Primer (Second Edition). Cambridge, Massachusetts, London, England, MIT Press (2008)
- Nikov, A., Jovanovik, M., Stojanov, R., Petkovski, M., Trajanov, D.: Use of Semantic Web Technologies for Meeting Management Software Development. In: ICT Innovations (2009)
- Bakraceski, G., Stojanov, R., Jovanovik, M., Trajanov, D.: A Knowledgebase Editor Using Semantic Web Technologies. In: 7th International Conference for Informatics and Information Technology (2010)
- Apostolski, V., Stojanov, R., Jovanovik, M., Trajanov, D.: Use of Semantic Web technologies in Knowledge Management. In: 7th International Conference for Informatics and Information Technology (2010)
- Zdraveski, V., Jovanovik, M., Stojanov, R., Trajanov, D.: HDL IP Cores Search Engine based on Semantic Web Technologies. In: ICT Innovations 2010, CCIS 83, pp. 306-315 (2011)
- Kulev, I., Janevska, E., Jovanovik, M., Stojanov, R., Trajanov, D.: Text Classification Using Semantic Networks. In: 8th International Conference for Informatics and Information Technology (2011)
- Böhm, C., Naumann, F., Freitag, M., George, S., Höfler, N., Köppelmann, M., Lehmann, C., Mascher, A., Schmidt, T.: Linking Open Government Data: What Journalists Wish They Had Known. In: 6th International Conference on Semantic Systems (2010)
- Cyganiak, R., Maali, F., Peristeras, V.: Self-Service Linked Government Data with dcat and Gridworks. In: 6th International Conference on Semantic Systems (2010)
- Berners-Lee, T.: Putting government data online. http://www.w3.org/DesignIssues/GovData.html
- 20. Alani, H., Dupplaw, D., Sheridan, J., O'Hara, K., Darlington, J., Shadbolt, N., Tullo, C.: Unlocking the potential of public sector information with semantic web technology. In: 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference (2007)
- 21. Data.gov.uk Opening up government, http://data.gov.uk/
- 22. Data.Gov An Official Web Site of the United States Government, http://www.data.gov/
- 23. Digitaliser, http://digitaliser.dk
- 24. Suomi.fi One address for citizens services, http://www.suomi.fi/suomifi/english/index.html
- 25. Aporta, Spain's Open Data Portal, http://www.aporta.es
- 26. London.gov.uk Greater London Authority, http://london.gov.uk/
- 27. Zaragoza Open Data Portal, http://www.zaragoza.es
- 28. Piedmont Open Data Portal, http://dati.piemonte.it
- 29. Helsinki Open Data Portal, http://www.hel.fi
- 30. The World Bank Working for a World Free of Poverty, http://data.worldbank.org/
- 31. dotNetRDF, An Open Source C# .Net Library for RDF, http://www.dotnetrdf.org/
- 32. Resource Description Framework, http://www.w3.org/TR/rdf-concepts
- RDF Vocabulary Description Language 1.0: RDF Schema, http://www.w3.org/TR/rdfschema/
- OpenLink Software. OpenLink Virtuoso Universal Server: Documentation, http://docs.openlinksw.com/virtuoso/
- 35. OWL Web Ontology Language Overview, http://www.w3.org/TR/owl-features/
- 36. GRDDL, http://www.w3.org/2004/01/rdxh/spec
- 37. RDFa, http://www.w3.org/TR/xhtml-rdfa-primer/
- 38. W3C Validation Service. RDF Validator, http://www.w3.org/RDF/Validator/
- 39. ReDeFer, http://rhizomik.net/html/redefer/
- 40. Visual Data Web. RelFinder, http://www.visualdataweb.org/relfinder.php

- 41. Jena A Semantic Web Framework for Java, http://jena.apache.org/
- 42. Protégé, http://protege.stanford.edu/
- 43. JRDF An RDF Library in Java, http://jrdf.sourceforge.net/
- 44. The Friend of a Friend (FOAF) project, www.foaf-project.org/
- 45. SPARQL Query Language for RDF, http://www.w3.org/TR/rdf-sparql-query/
- 46. YSlow Yahoo! Developer Network, http://developer.yahoo.com/yslow/

Enhancing Scholarly communication results and search experience by interlacing relevant Scientific Repositories

Arben Hajra¹, Vladimir Radevski¹, Atif Latif²

¹South East European University, Faculty of Contemporary Sciences and Technologies, 1200 Tetovo, Republic of Macedonia {a.hajra, v.radevski}@seeu.edu.mk
²German National Library of Economics (ZBW) Leibniz Information Center for Economics, 24105 Kiel, Germany {a.latif}@zbw.eu

Abstract. Semantic Web is an attempt to extend the current web thus the data to be machine understandable and independent from platforms. In this way machines can understand the meaning of the data and automatically accomplish tasks. The concept of Linked Data gives a set of best practices for publishing and interlinking structured data on the Web, between different data sources, effectively allowing data in one data source to be linked to data in another source. Following this, the goal of this research is the usage of LOD for enriching libraries for additional recommendations. The current static metadata structure of Digital Libraries does not offer the most relevant, up to date and cross domain literature, which leads to limitations in the literature search and information retrieval. For increasing these benefits and enhancing search experience, it is necessary to interlace repositories of different domains inside the LOD cloud or with resources in Web 2.0.

Keywords: semantic web, digital libraries, interoperability, linked open data

1 Introduction

The Semantic Web is an initiative that aims at improving the current state of the World Wide Web [3] – not by offering an alternative for replacing the current Web but by offering an attempt to extend it. The main aim is achieving that the data on the Web to become machine understandable information, independently of platforms and other boundaries. The Semantic Web provides the technologies and standards that are needed to add machine-understandable meanings to the current Web, so that computers can understand the Web documents and therefore can automatically accomplish tasks [4].

Since the current Web is designed to be used by the humans, there is a proposed solution from W3C for describing the data in the Web - metadata to enable automated processing of the related Web resources [4].

The implementation of semantic technologies and the interoperability between different linked data sources can be used for enriching digital libraries with additional recommendations. The role of the current digital libraries is more than evident; however there are several directions where they lack to provide the needed service. The benefits of interoperability between different data repositories give a new vision about scholar community, scientific findings, knowledge representation and retrieving.

2 Digital Libraries

Digital Libraries (DL) represent an important place for storing, organizing and retrieving information. They can be categorized as a primary source from where scholars are provided with resources. At present there is a large number of the DL, in different levels, as national, international, institutional, thematic, etc. Their usages bring a huge avail for the community and for scholars especially, with their global and simple accessibility.

However, in some cases getting the most relevant and qualitative resource using a digital library can be a challenge. Searching in a specific DL will result a list of resources, indexed and cataloged using a static metadata structure in that library. At the actual stage the scientific publications are being catalogued by trained professionals in a field or librarians. This process leads to limitations in the literature search to a specific field – correspondingly catalogued and indexed. According to [25], DL are noted as Monolithic Systems, where metadata describe the data rather than uses.

Another issue that has been a central concert of DL from the beginning of their creation is the interoperability between them. As it is mentioned in [27] researchers have been struggling with interoperability, as one of the main feature for achieving better recommendation results from digital libraries.

3 Problem Statement

In the previous section we emphasized the importance of current digital libraries as primary source for scholarly communications. However, there are present several challenges in different fields that can prevent the scholar for getting the best of them.

The static metadata structure, where metadata are describing the data instead of the usage of that data, can be one of the limits for having the most relevant and up-to-date information. For a given search in a digital library, the result will be the information for the indexed resources described with a specific metadata structure. No additional information on the relevance, usability and quality of the retrieved resources will be available. Even in several digital libraries, as ACM DL, where there is some sorting criteria for this purpose still there is a lack of community presence in the content evaluation. The number of views, downloads, cites, tags, shares, bookmarks, etc., can be an important indicator for the relevance of a source and also should be taken into consideration.

Actually, if a person is searching at Digital Library about "Global Warming", will get a list of indexed publications catalogued and described by a static metadata structure. For additional recourses should manually search at other DLs or looking different resources as blogs or wikis. This example show that, starting from one DL it is impossible or at least very difficult to cross the boundaries by spreading your search in other resources. If we refer to the current structure, APIs from several services can contribute in this direction, but again we are with fixes sources where changes and new updates are limited.

4 Linked Open Data

Linked Data has been introduced and conceptualized in 2006 by Tim Berners-Lee as a set of best practices for publishing and interlinking structured data on the Web [1] and to provide broad guidance [2]. Linked Data is about employing the Resource Description Framework (RDF) and the Hypertext Transfer Protocol (HTTP) to publish structured data on the Web and to connect data between different data sources, effectively allowing data in one data source to be linked to data in another data source [2],[5],[6]. Actually on the Web are different and interesting open data sets, the figure 1 give a part of the Linked Open Data Cloud Diagram.

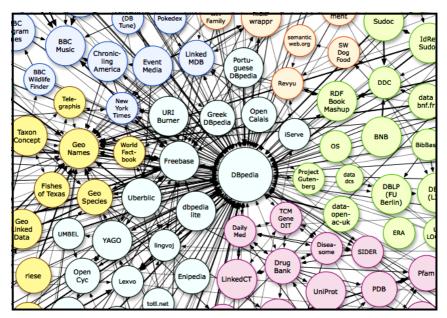


Fig. 1. Part of the Linked Open Data Cloud Diagram

Various datasets are published on the web, under specific license or as public. Among others we can list Wikipedia, DBLP bibliography, Wikibooks, Geonames, FOAF, MusicBrainz, WordNet, German National Library and many more which are published under Creative Commons or Talis licenses [8]. Also, Google, Yahoo, Microsoft (at schema.org¹), Facebook and others has agreed to publish structured data on the web.

The increased number datasets starting from June 2007 show the importance and the interest for publishing linked data. In September 2011, 295 data sets are counted in LOD cloud, consisting of over 31.6 billion RDF triples, which are interlinked by around 504 million RDF links [9]. The most of datasets from this cloud belong to the Publication domain, Government, Media, Geographic, etc.

Comprehensive Knowledge Archive Network (CKAN) powers a number of data catalogues on the Internet, where each dataset contains a description of the data and other useful information [10]. Everyone can create a dataset for specific field or purpose so; DataHub [10] gives a visual representation of these datasets according to specific criteria, as total number, weekly revised, top rated, format etc. Figure 2 show the current number of datasets.

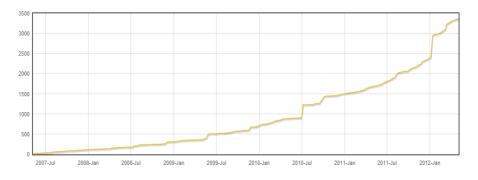


Fig. 2. The total number of datasets [10].

From Table 1 can be noted the largest groups of datasets based on datahub statistics that consists the LOD cloud. The most important domain in this case is about publications which include library and museum data.

 Table 1. The number of RDF links refers to out-going links that are set from data sources within a domain to other data sources [10].

Domain	Datasets	Triples	(Out-)Links
Media	25	1,841,852,061	50,440,705
Geographic	31	6,145,532,484	35,812,328
Government	49	13,315,009,400	19,343,519
Publications	87	2,950,720,693	139,925,218
Cross-domain	41	4,184,635,715	63,183,065
Life sciences	41	3,036,336,004	191,844,090
User-generated content	20	134,127,413	3,449,143

¹ Provides a collection of schemas, i.e., html tags, to markup their pages in ways recognized by major search providers, including Bing, Google, Yahoo! and Yandex

Bellow we list some Libraries offered as Linked Open Data, that can be noted inside the LOD could and can be of interest for our scenarios:

German National Library (DNB), Swedish National Library (LIBRIS), British National Bibliography (BNB), Hungarian National Library (NSZL), Europeana Digital Library, Library of Congress Bibliographic Records, ZBW - STW Thesaurus for Economics, DBLP Bibliography Database, Association for Computing Machinery (ACM), etc.

It is interesting to mention that the growth of Library Linked data for one year, 2009-2010, has been one thousand percent [24], which show the interest for publishing this type of data. In this context the importance and the role of Open Linked Data as a way to link and query data from different resources is more than evident.

5 Proposed Solution

For enhancing search experience in scholarly communication and getting a solution for the problems stated above, the proposed research track is in the domain of Semantic Web. The search results would be an outcome of a system that would overcome the limitation of catalogues and indexation in a particular scientific domain and will take into consideration other scientific domains. The results should offer answers of the best ways to proceed in order to offer shared views on relevant literature - very often multidisciplinary by character, and thus allow discovery experience as well as search over variety of scientific areas and domains. This will provide a shift to the scholarly communication to a new level of connectivity among researchers fostering the dissemination, sharing and discovery processes.

The usage and interoperability with scientific resources offered as Linked Open Data can be a good track for reaching the goal; to enrich the library with recommendation from LOD. In this way, the aim would be to deliver web services which will automatically locate, identify and link up related resources from LOD.

A solution and implementation of the usage of LOD for discovering and constructing author's profiles shows the advantages of using this type of data [11]. Instead of looking author's websites and collecting information from several pages manually, the system can integrate the most relevant data about specific author using LOD repositories and represent them in an appropriate form.

In a case where the results getting from interoperability with LOD do not satisfy the needs, we can refer to Web 2.0 resources as scholarly digitally maintained libraries. Actually there are several socially maintained scholarly libraries, based on the folksonomy approach, as a social resource sharing systems. It means that a resource at Web 2.0 can be tagged, commented, ranked, thus in an "informal" way gets a kind of meaning [19], [20], [21]. Therefore, a specific publication from these repositories now can be judged by popularity (e.g. most visited, most downloaded), comments, ranking, etc. The challenge in is this situation is to discover scholarly resources from socially maintained library services, as Bibsonomy², Mendeley³,

² http://www.bibsonomy.org/

³ http://www.mendeley.com/

CiteULike⁴, etc. Thus shared views on relevant literature and more socially influenced approaches can enrich the scholars' discovery experience.

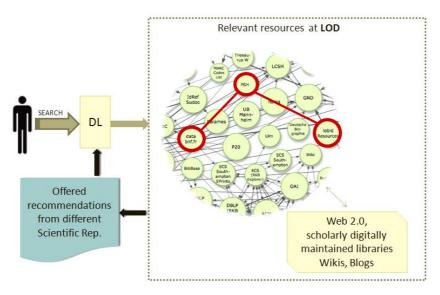


Fig. 3. Enrich the library with recommendations from LOD or Web 2.0

Unlike the current search, when a person would look at the DL about "Global Warming" the system automatically will offer a list of resources not only from the DL where the search is performed, but will merge and integrate data from several fields that in a way are related with Global Warning. So, the system can generalize or specify the keyword and the offered publications would be from repositories like, meteorology, ecology, economy, medicine, etc., that in a way meet the search criteria. At the same time the system would check the quality and the relevance of the retrieved data. Therefore, a DL can be a start point for searching other sources as a global database.

6 Research Track

The large number of scientific resources offered as Linked Open Data and the interoperability between them can be a huge benefit for scholars and for the community in general.

Thus the goal of the research is to enrich the library with recommendations from LOD or Web 2.0. In this way, would be crossed the scientific boundaries of digital libraries by ensuring interoperability of library data (Conventional and LOD) for enhancing search experience, by linking and querying data from different repositories.

⁴ http://www.citeulike.org/

This can result in recommendation generation, more qualitative content enrichment, up to date and accurate literature search, etc.

As it is mentioned at [22], [24] some major research challenges should be taken into consideration for achieving this goal, as:

- Web of Data Access and identifying relevant resources
- Vocabulary Mapping and Identity Resolution
- Integrating with Conventional Library systems
- Quality Evaluation or User Evaluation

6.1 Identifying relevant resources

To identify and link up relevant resources related to library domain can present a real challenge. It has been stated [24] that is very difficult to discover what is in the data space and how does the content change over time. For our research it will be of significant importance the fact that we will focus to the domain of the library datasets. The first attempts can be done using and interlacing datasets as DBpedia, DBLP, Freebase as more general and other specific datasets, which are mentioned in section four. The current practice of accessing these data has been through crawling, SPARQL endpoints, RDF data dumps or using APIs offered by some LD search Engines, like Sindice⁵, Watsoon⁶ or Swoogle⁷.

6.2 Mapping and Identity Resolution

Heterogeneity between different RDF vocabularies for representing the same type of information, present the need for translating terms into a target schema. Therefore, for achieving the interoperability the diversity of the schemas for describing metadata such as Dublin Core, FOAF and a variety of vocabularies accordingly, need to be mapped and aligned. This may lead to wider interoperability by linking to existing vocabularies, schemas and taxonomies [22].

Identity resolution heuristics should be applied for discovering the things in different repositories in LOD. Several repositories may have *owl:sameAs* links for referring at the same things in other data sources. However, much work need to be done for the case where does not exists such links. Data provenance is an issue, very important for judging the quality of the data by analyzing their origin. The current state of LOD cloud, show that 47.13% of library datasets provide provenance.

According to [24], [26], for large-scale integrations, with thousands of data sources it is impossible to model a unifying integration schema upfront. Thus much more effort should be done on generating and publishing mappings on the web, and the same can be discovered and used for other data integrations. In current state data

⁵ http://sindice.com/

⁶ http://kmi-web05.open.ac.uk/WatsonWUI/

⁷ http://swoogle.umbc.edu/

⁸ http://www4.wiwiss.fu-berlin.de/lodcloud/state/#provenance

integration is achieved in a pay-as-you-go manner, because only a small number of mappings have been added to the system.

6.3 Quality evaluation

Quality of the integrated data is very important. The process of data integration from different repositories can produce redundancy, stale, outdated or simply wrong data. In this context the data quality assessment becomes crucial for which the user evaluation can be very helpful for the process in general.

7 State of the Art

Identity resolution and vocabulary mapping are one of the key points in the integration process. In this context, SameAs⁹ [14][15] from RKB Explorer¹⁰ provides a service to find equivalent URIs, thus related data about a given resource from different sources [13].

<sameas> interlinking the Web of Data</sameas>		
The Web of Data has many equivalent URIs. This service helps you to find co-references	<sameas> Enter a Linked Data URI</sameas>	٩
between different data sets. Enter a known URI, or use Sindice to search first.	SEEU SEEU	٩
Search results from Sindice, with co-references applied Services [Other] SBS SV01 SeeU for VOCALOD3AE interview 1. http://www.orencik.net/forum/Tread-oth refraint 0. jpm - test Service [Other] SBS SV01 SeeU for VOCALOD3AE interview 1. http://www.orencik.net/forum/Tread-oth refraint 0. jpm - test Service "21.12.2011 TETOVE - Seeu campus" Q interview 0. jpm - test	ASSIGN [™] Q, ersbs-sv01-seeu-for-vocaloi⊥ign-77187 ASSIGN [™] Q, ersbs-sv01-seeu-for-vocaloi⊥ign-77192	

Fig. 4. The sameAs.org look-up service

The generic services of linked data search engines such as Sindice, Watson or Swongle allow searching by keywords and URIs for querying the aggregated data from LOD cloud [15][16].

About the quality of data, *Falcons*¹¹ search engines is a good example that enables the user to filter search results and therefore limit the results to show, also offer to users to eliminate low quality data. The linked data browser, *Sig.ma*¹², besides allowing to users to navigate at the merged data sources by following RDF links also

⁹ http://www.sameas.org

¹⁰ http://www.rkbexplorer.com/data/

¹¹ http://iws.seu.edu.cn/services/falcons/

¹² http://sig.ma/

apply vocabulary mapping. While, *Marble*¹³ track the provenance of data while merging data about the same thing from different sources.

There are several scopes where interlinking between resources can be productive. An example of this one is the ELIXIR¹⁴ project which purpose is to create and maintain a sustainable infrastructure for biomedical research community in Europe [17]. An interesting tool for performing a search in digital document repositories in the biomedicine domain Kleio [18], is a combination of textual and metadata search. The semantic metadata results are generated from the text mining processes.

For increasing the interoperability of library data on the Web, very active is W3C *Library Linked Data Incubator Group*¹⁵, by exploring existing building blocks of librarianship, such as metadata models, metadata schemas, standards and protocols of current digital libraries.

There are several successful implementations in this scope, as *JeromeDL*¹⁶. It is a Social Semantic Digital Library which allows institutions to publish documents on the Web where every library user can bookmark interesting books, articles or other materials in semantically annotated. Users can allow others to see their bookmarks and annotations and share their knowledge within a social network.

*VIVO search*¹⁷ is an example, for storing and exposing data as RDF, where the information can be displayed or delivered to other systems as RDF. Thus the metadata would be aggregated and integrated into the Linked Open Data cloud. Currently has a demonstration of multi-institutional search between several universities as, Cornell, Harvard, Indiana and Florida.

8 Conclusion

According to the use cases where Semantic Web has been deployed [12], the main benefits of SW are in data integration and search improvement.

The interoperability between the scientific resources, which nowadays appear in a large number, can offer a huge benefit for improving the search experience of digital libraries. For increasing these benefits, it is necessary to interlace repositories of different domains in general, inside the LOD cloud or with resources in Web 2.0.

The research track of our interest will focus on the possibility of automatic location, identification and linking up related resources from Linked Open Data with arbitrary library, correspondingly indexation and categorizations platforms.

References

1. Berners-Lee, T.: Linked Data-Design Issues, http://www.w3.org/ DesignIssues/ LinkedData.html, (2006)

¹³ http://www5.wiwiss.fu-berlin.de/marbles/

¹⁴ http://www.elixir-europe.org

¹⁵ http://www.w3.org/2005/Incubator/lld

¹⁶ http://www.jeromedl.org

¹⁷ http://www.vivoweb.org

- Bizer, C., Heath, T, Idehen, K., Berners-Lee, T.: Linked Data on the Web (LDOW2008), WWW 2008, April 21–25, Beijing, China, ACM 978-1-60558-085-2/08/04, (2008)
- 3. Antoniou , G., Hemrelen, F.: A Semantic Web Primer, The MIT Press, ISBN 978-0-262-01242-3, (2008)
- 4. Yu, L.: A developer's guide to the Semantic Web, Springer, ISBN 978-3-642-15969-5, (2011)
- 5. Hebeler, J., Fisher, M., Blace, R., Perez-Lopez, A.: Semantic Web Programming, Wiley Publishing, Inc., ISBN 978-0-470-41801-7, (2009)
- W3C Rec, Resource Description Framework (RDF) Model and Syntax Specification, Editors: Ora Lassila, Ralph Swick, http://www.w3.org/TR/1999/PR-rdf-syntax-19990105, (1999)
- 7. W3C, Semantic Web Standard, http://www.w3.org/standards/semanticweb/, (Accessed April 6th, 2012)
- 8. DuCharme, B.: Learning SPARQL, Reilly Media, Inc., ISBN 978-1-449-30659-5, (2011)
- LinkingOpenData, W3C SWEO Community Project, http://www.w3.org/wiki/SweoIG/ TaskForces/CommunityProjects/ (Accessed April 8th 2012)
- 10. CKAN-Data hub, Statistics, http://thedatahub.org/stats (Accessed April 8th 2012)
- Latif, A., Afzal, M. T., Helic, D., Tochtermann, K., Maurer, A. H.:Discovery and Construction of Authors' Profile from Linked Data A case study for Open Digital Journal, In: LDOW, Vol. 628CEUR-WS.org (2010)
- 12. W3C.org use cases and projects, http://www.w3.org/2001/sw/sweo/public/UseCases/, (Accessed April 4th 2012)
- Latif, A., Afzal, M., Hoefler, P., Saeed, A., Tochtermann, K.: Turning keywords into URIs: simplified user interfaces for exploring linked data. In Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human (ICIS '09). ACM, New York, NY, USA, 76-81. DOI=10.1145/1655925.1655939 http://doi.acm.org/10.1145/1655925.1655939 (2009)
- 14. Glaser, H., Millard, I., RKB Explorer: Application and Infrastructure, In: Semantic Web Challenge, Vol. 295CEUR-WS.org (2007)
- 15. Hausenblas, M., Utilising Linked Open Data in Applications, WIMS'11, Sogndal, Norway, ACM 978-1-4503-0148-0/11/05 (2011)
- 16. Oren, E., Delbru, R., Catasta, M., Cyganiak, R., Stenzhorn, H., Tummarello, G.: Sindice.com: a document-oriented lookup index for open linked data. Int. J. Metadata Semant. Ontologies 3, 1 (November 2008), 37-52. DOI=10.1504/IJMSO.2008.021204 http://dx.doi.org/10.1504/IJMSO.2008.021204 (2008)
- 17. Repanas, K., Rebholz-Schuhmann, D., Valencia, A.: Elixir WP8: "Scientific Literature Interdisciplinary Interactions", Final Report (2008)
- Nobata, C., Sasaki, Y., Okazaki, N., Rupp, C. J., Tsujii, J., Ananiadou, S.: Semantic Search on Digital Document Repositories based on Text Mining Results, International Conferences on Digital Libraries and the Semantic Web 2009 (ICSD2009), page 34-38, (2009)
- Anderson, P.: What is Web 2.0?Ideas, technologies and implications for education, JISC Technology & Standards Watch, (2007)
- 20. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: BibSonomy: A Social Bookmark and Publication Sharing System, Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures, (2006)
- 21. Vander Wal, T.: Folksonomy, Blog post., http://vanderwal.net/ folksonomy.html (2007)
- 22. Yu, H.Q., Dietze, S., Li, N., Pedrinaci, C., Taibi, D., Dovrolis, N., Stefanut, T., Kaldoudi, E., Dominque, J.: "A Linked Data-driven & Service-oriented Architecture for Sharing Educational Resources", In the Proceedings of the Linked Learning 2011: 1st International Workshop on eLearning Approaches for the Linked Data Age, Herakleion, Greece, May 29-30, (2011)

- 23. Peristeras, V.: From Open Linked to Gorverment Data, Interoperability Solutions for Eruropean Public Administration, ppt presentation (2010)
- 24. Heath, T., Bizer, C.: Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool. (2011)
- 25. Borgman, Ch.: Personal digital libraries: Creating individual spaces for innovation, NSF post-DL workshop paper, June 4, (2003)
- 26. Halevy, Y.A., Norvig, P., Pereira, F.: The unreasonable effectiveness of data. IEEE Intelligent Systems, 24(2):8–12, 2009. DOI: 10.1109/MIS.2009.36 106 (2009)
- 27. Paepcke, A., Chang, Ch. K., Winograd, T., Molina, H. G.: Interoperability for digital libraries worldwide. Commun. ACM 41, 4 (April 1998), 33-42. DOI = 10.1145/273035.273044 http://doi.acm.org/ 10.1145/273035.273044 (1998)

Relational Databases in Data Mining on Grid

Ğorgi Kakaševski¹, Anastas Mišev²

 ¹ Faculty of Information-Communication Technologies, FON University, bul. Vojvodina b.b. 1000 Skopje, R. Macedonia gorgik@fon.edu.mk
 ² Institute of Informatics, Faculty of Sciences and Mathematics, Ss. Cyril and Methodius University, 1000 Skopje, Republic of Macedonia anastas@ii.edu.mk

Abstract. Data mining and statistical analysis algorithms are a set of heuristics and calculations that creates a data mining model from data. The mining model that an algorithm creates from data can take various forms: a set of clusters, a decision tree, a mathematical model or a set of rules. With producing more and more digital data, placed on several locations, need of distributed data mining and statistical analysis algorithm increases. Grid computing provides a platform that can be used for this purpose, utilizing both, the computing and data resources. While most of Grid data is file based, the need for access to structured relational data is increasing. This paper presents specifications for Data Access and Integration on Grid, and OGSA-DAI which is practical implementation of these specifications. As a conclusion of this paper, we propose several visions for future work.

Keywords: Grid computing, heterogeneous databases, distributed query processing, OGSA-DAI.

1 Introduction

E-Science is computationally intensive science that is carried out in highly distributed network environments, or science that uses immense data sets [1]. Digital data play major role in e-Science applications, from small projects [2] to The Large Hadron Collider [3].

The computational needs of today's data mining algorithms are constantly increasing. The Grid provides both computational and storage resources to the scientific applications [4]. Via dynamic and distributed Virtual Organizations, Grid computing technologies provides approach to use geographically distributed heterogeneous large-scale data resources. The key characteristics distinguishing Grid computing over the conventional technologies are: it is flexible, autonomous, dynamic and service-oriented. In this are included systematic and effective access and interaction of distributed data resources.

Large-scale data is the basis of data mining applications on Grid. Recently, various systems and approaches to grid-based data mining have been reported in the literature: GridMiner [5], DataMiningGrid [6], Knowledge Grid (K-Grid) [7],

Anteater [8], Adam [9] etc. This systems are service oriented according to the Grid standards. They support data mining and online-analytical processing in distributed computing environments and implements some common data mining algorithms, including parallel versions, and several forms of knowledge representation. All of this systems use structured or unstructured data, but before data processing, they export data into files. They don't have incorporated distributed database management system. This creates gap between data and processing elements of complex data mining workflow.

Because of mainly file-based organization of Grid data, there is a need how to use data which are originally organized as shared and structured collections, stored in relational databases, or in some cases in structured documents or in assemblies of binary files. These requirements of Grid applications require developers of Grid systems to provide practical implementations of "relational Grid databases" (or data access integration [10]). Furthermore, data mining algorithms have several stages, from pre-processing to generating graphical representation, and these stages can be executed in some form of workflow [11]. For this reasons relational Grid database system should be easily extendable and adoptable for different purposes.

In this paper we discuss standards for data access and integration and their practical implementation which can incorporate relational (and other types of) databases in new Grid technologies. Our goal is to show that these distributed database systems, can be source or base of data mining system on Grid and such a system has relational database more closely to the Grid – instead of exporting files from database. As result of this work we give several visions for future work in this field.

The rest of the paper is organized as follow. In Section 2 we describe standards for data access and integration which are key concepts for practical implementations of relational database middleware for Grid. We described OGSA-DAI, a practical release of WS-DAI specifications in Section 3. These database systems can be used as base of data mining systems on Grid which are described in vision for future work in Section 4. We conclude paper in Section 5.

2 Standards for Data Access and Integration

One of the main characteristics of the Grid is it heterogeneity, which means that Grid is made of different type of resources with different characteristics, and as whole must act as one system. For example, on the Grid we can find two clusters with different type of processors or installed software, even on the same cluster we can find machines with different resources. This property is true for databases. There are several types of databases which can be used in applications: relational, XML, even CSV files or just files. But The Grid users (applications, clients) must see data as one integrated part, no matter where they are located or which their physical form is.

Another key moment in organizing databases on Grid is the service-oriented architecture of Grid and Grid services, defined by Open Grid Service Architecture (OGSA) [12]. OGSA has a one of the central roles in Grid computing, because it simplifies the development of secure, robust systems and enable the creation of

interoperable, portable, and reusable components. The main idea of transient Grid services is to extend well-defined Web service standards (WSDL, XML schemas, SOAP, etc.), to support basic Grid behaviors. These standards are proposed by Open Grid Forum and this organization have key role in development of new Grid technologies.

Open Grid Forum (OGF) [13], earlier Global Grid Forum, is an open community committed to driving the rapid evolution and adoption of applied distributed computing. Applied distributed computing is critical to developing new, innovative and scalable applications and infrastructures that are essential to productivity in the enterprise and within the science community. OGF accomplishes its work through open forums that build the community, explore trends, share best practices and consolidate these best practices into standards.

Grid computing provides service-oriented middleware that supports secure and managed sharing of networked computational resources. Grid enables activities which involve collaboration on different levels across many domains. For efficient functioning of such kind of distributed system OGF proposed shared abstractions and consistent interfaces. The OGF is the principal standards body for Grid computing and mainly develop web service standards which are relatively independently or as part of a wider service-based architecture.

OGF issued the Web Service Data Access and Integration (WS-DAI) family of specifications [14]. This specifications defines web service interfaces to data resources, such as relational or XML databases. On Fig. 1 are shown main components of WS-DAI specifications.

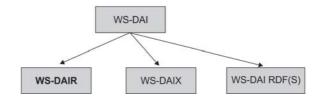


Fig. 1. WS-DAI family of specification

In particular, the WS-DAI specifications are part of a wide ranging activity to develop the Open Grid Services Architecture (OGSA) within the OGF. The specification WS-DAI is on the top level and it defines the core data access services which are web service interfaces to data resources.

WS-DAI specifications have five main components: interfaces, data services, properties, messages and data resources. Interfaces are used to interact with data services. Data services are web services that implements interfaces to access distributed data resources. Properties describe data services, messages are sent to data services and data resources are system that can act as a source of data. Data resources are already existent database management systems, for example, MySQL, and this system has no real connection with WS-DAI specifications.

On Fig. 2 is given the proposed messages flow between client and data stored on Grid, by Open Grid Forum. This scheme enables clients to access or modify the data and their organization through Grid data service.

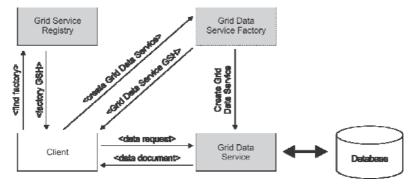


Fig. 2. Accessing and modifying data stored on Grid using a Grid Data Service

Three sub specifications extend the super specification WS-DAI. WS-DAIR is a specification for a collection of data access interfaces for relational data resources and extends the WS-DAI. It extends the properties defined in WS-DAI, instantiates the templates for direct and indirect data access, and defines interfaces for accessing the results of requests directed at a relational data service. XML realization is covered in WS-DAIX. Key features in WS-DAIX include: properties and operations for manipulating collections of documents, direct and indirect access using both XQuery and XPath, and data modification using XUpdate. WS-DAI RDF specification is work in progress and should provide access to RDF(S) data.

In order to implement these specifications on real system, on Fig. 3 is given proposed architecture of Grid database system [16]. During the fact that these traditional DBMSs exist for many years and are well developed and tested, in this case can be used some of the existing. On top of this system Grid Data Services must be implemented, as part of the middleware that is responsible for communications among heterogeneous distributed database.



Fig. 3. Architecture of Grid Database System

Grid Data Service Middleware is one kind of transparent wrapper for complex Grid database. Transparency refers to: location, name of database, distribution, replication,

ownership and cost, heterogeneity and schema. Clients exchange data with Grid database, even if they don't know where that data is stored or in which format. Important for client is that in Grid database must be implemented all (or common) functionalities of regular, mainly relational, database management systems and in natural way. That ensures that clients (users, software developers) can easily use Grid database and easily adapt their applications with new kind of database.

As a conclusion, a heterogeneous distributed database system (or simply Grid database) should provide service-based access to existing relational database management systems that are located on Grid nodes. This database will be some form of virtual database and clients will access to data on transparent way no mater of physical organization of data [16].

3 Practical implementation of WS-DAI

To be accepted one OGF standard it must have at least two independent implementations. In WS-DAI and WS-DAIR Implementations - Experimental Document [17] is given reports on the interoperability testing results arising from two implementations of WS-DAIR (which also implement WS-DAI). By interoperability is mean interoperability of the clients accessing the DAI defined services, i.e. not inter-service interoperability. To be more specific, the actual SOAP messages sent by a client, in this case a third party application, must produce the same behavior and results for each of the DAIS compliant services in order for those services to be defined as interoperable.

The implementations that participated in this exercise were:

- AMGA WS-DAIR implementation [18].
- OGSA-DAI WS-DAIR implementation [19].

AMGA WS-DAIR is developed as a part of gLite middleware. Its implementation uses gSoap & C++, supports the following underlying database management systems: PostgreSQL, MySQL, SQLite and Oracle. Supported languages for querying data are SQL-92 entry level "direct data statement" and the AMGA metadata language, but it doesn't support stored procedures. It supported WebRowSet dataset and this security features: SSL, GSI, VOMS, permission, and ACL.

OGSA-DAI stands for Open Grid Services Architecture - Data Access and Integration. It is more popular and used by many projects [20][21], and for that reason we analyze in more details. It is a Java open source project and it is a one of the most comprehensive heterogeneous database system which can be used in Grid.

OGSA-DAI make the DBMSs, which can be heterogeneous and distributed on Grid, transparent to the users [22]. It hides the underlying drivers that are needed to communicate with the database management and as name tells, OGSA-DAI provide service-oriented interface to the clients. This process naturally will add some overhead on top of the communication itself but it is not that significant.

It supports several kinds of databases: relational (Oracle, DB2, SQL server, MySQL, Postgress), XML databases, like xindice and eXist, and files. There is an opportunity for programmers to develop new drivers for DBMS which are not

supported in original version. Acting in the top of local DMBS that are located on different sites, OGSA DAI provides uniform Grid database.

Today, OGSA-DAI is a part of many scientific projects that are using it as a Grid database (i.e. ADMIRE1, BIRN2, GEO Grid [14], MESSAGE [15], BEinGRID5, LaQuAT6, Database Grid7, mantisGRID etc.). List of most important projects can be found in [23].

In rest of this chapter we describe architecture of OGSA-DAI and its part called OGSA-Distributed Query Processing (OGSA-DQP).

3.1 Architecture of OGSA-DAI

Architecture of OGSA-DAI is in compliance with needs described in Section 2. In OGSA-DAI, data resources represent externally managed database system and data service provide interface to access to this data resources.

Activities are units of work and they perform well defined data-related task over a specific resource. Main activities are organized in three types: Statement Activity, Translation Activity and Delivery Activity (Table 1). Group of activities form query-transform-deliver workflow, and users interact with OGSA-DAI services by submitting workflows to Data Request Execution Service (DRES) [24]. On Fig. 4 a) is shown example of OGSA-DAI workflow with three activities.

Type of activity	Example
Relational Database	SQLQuery, SQLUpdate, SQLBulkLoadTuple,
	SQLParameterisedQuery, ExtractTableSchema.
XML Database	XPathQuery, XQuery, XUpdate, AddDocument, GetDocuments,
	RemoveDocuments
Transform	ByteArraysToTuple, CSVToTuple, XSLTransform,
	StringTokenize
Delivery	DeliverToDataSink, DeliverToFTP, DeliverToSMTP,
	DeliverToGFTPExtended, DeliverToSession, ObtainFromFTP
DQP	CreateDQPResource, ExtractTableSchema (DQP), SQLQuery
	(DQP)

Table 1. List of main activities in OGSA-DAI

On Fig. 4 b) is shown the interaction between client and data through DRES.

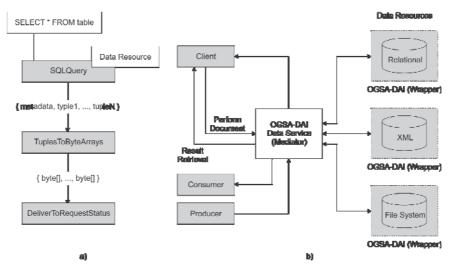


Fig. 4. a) Example of OGSA-DAI query-transform-deliver workflow. b) Interaction between client and data.

The first step in this process is sending perform document from client to DRES. This document is XML based document which include sequence of activities. This activities are well defined tasks which will be completed on server side and this activities are connected one another to form a workflow where output from one activity is an input to another. Finally, result data can be delivered back to the client, or to third party (consumer).

3.2 OGSA-DQP

Because data is distributed across various resources, distributed database management systems have a need for effective distributed query processing. This component should allow optimization of SQL queries in system. OGSA-DAI has service-based OGSA-DQP [25] component which allow distributed query execution. It enables distributed queries over relational data resources. Although the core data service in OGSA-DAI provide abstraction for accessing individual data resources on the Grid, they do not address challenges associated with integrating data from multiple resources.

DQP coordinator service and DQP evaluator service are two services that extend OGSA-DAI. The first one have a DQPQueryStatement activity which compiles, optimizes and schedules SQL queries for execution. The DQP evaluator service is capable of evaluating a query fragment provided by the coordinator. Evaluator communicates with OGSA-DAI resources, and to provide benefits of parallelism, multiple evaluators are used.

On the beginning, OGSA-DQP obtains the metadata that is needs to compile, optimize, partition and schedule distributed query plans over the multiple execution nodes (evaluators). After the submitting of query, the process is divided in two steps:

(i) the query plan is produced and partitioned, and each partition is sent to the relevant evaluator; (ii) the evaluators retrieve data from OGSA-DAI data services and invoke any analysis services required to evaluate query. Evaluator is exploiting partitioned parallelism both for query internal operations (such as join) and for external web service calls.

In the newest version (OGSA-DAI 4.1), there are some changes and improvements [24]. Major change is that DQP evaluator service is removed, an all of its functionality is now provided by new activities in OGSA-DAI. New version support more SQL operations, improve extensibility, introduced new way of configuring. However there are still shortcomings and some of them are mentioned in Section 4.

4 Visions for Future Work

Although today we have several implementations of Grid databases, this area is relatively new and there is a need for future work. Also, there is a lot for done in the field of Grid data mining.

Many of the Grid data mining systems [5][6][9] are not database oriented and they are independent systems which should be ported on Grid. Several disadvantages occur here: systems should be ported on several Grid sites, they rarely use relational database, data are separated from algorithms and data are mainly file based. These create unnatural flow of data which is very important in process of data mining.

To have more efficient data organization in data mining applications on Grid, we propose three main steps which should be analyzed in future. Our vision for future work includes:

analyzing standards for Data Access and Integration,

• upgrading heterogeneous and distributed database which is compliance with OGF WS-DAI standards and have good compatibility with Grid,

upgrading database system with data mining needs.

OGF is still working on WS-DAI specifications. People who are related in this field could propose changes in this documents and its practical implementation. In manner of creation of WS-DAIR specification, which is derived from WS-DAI, some authors [26] propose WS-DAI Data Mining specification. OGF should assess whether such a specification is needed.

Second step refer to upgrading database systems. These systems are not ideal but good thing is that many of them have active developers, they are not forgotten projects and they are developed as open source projects. Database partitioning is one of the questions that should be solved in transparent way, moreover to make this process dynamic by Grid. Partitioning should split tables on different locations (horizontally or vertically), but gives a non partitioned database view for the client side when working with SQL commands. There are several benefits from database partitioning: manageability, scalability, performance, availability and security. There is a need for developing a new tools for monitoring, replica management, fault-tolerance, QoS [27] etc.

The projects mentioned before, as a good example of Grid database system, OGSA-DAI, can also be extended. The structure of this open source project is made

to be extensible in many different ways by: writing activities, writing data resources, writing application-specific presentation layers, to improve or add functionalities to security or database access. This enables third parties to continue the development of stabile relational database system with functionalities, instead of developing a new one. OGSA-DQP also has some limitations. Parallel query-processing optimizations utilizing additional evaluators are not yet supported. There is no GUI client (web based user interface). All the existing optimizers are essentially heuristic and do not make use of a cost model, join ordering makes use of very basic cardinality estimates, there is no dynamic fault tolerance, etc.

And our last proposed step is upgrading database systems with data mining needs. For example, according to the needs, the system can preprocess data such as data cleaning, distributing data at different sites, data transformation, data reduction, data density analysis, etc. These preprocessing activities can be executed in dependency of data amount and algorithms for preprocessing. If algorithm doesn't need a lot of computational power, then is much cheaper to execute preprocessing on the site where data are located. OGSA-DAI can be extended with different preprocessing activities which can be included later in modeling workflow. Because several data mining algorithms use similar preprocessing algorithms, here is possibility for building general preprocessing module which can be integrated within OGSA-DAI. In similar way OGSA-DAI can be extended with many other data mining activities. An attempt to do this is made in ADMIRE project [28], but this project is too related to OGSA-DAI instead of using other Grid architectures for computation. For example, to execute data mining tasks, OGSA-DAI must be installed on all data and computational nodes that is problem for existing Grid infrastructures.

5 Conclusion

We can highlight three conclusions as a results of presented work in this paper. First, Grid data mining applications have an increasing need of relational database systems. Combining Grid and database technologies is an essential approach to meet the requirements of large-scale Gird applications. Almost every data mining application running on Grid has many requirements for access to structured relational data, and Grid as a platform with their resources can provide many benefits for such kind of database system.

Our second conclusion is that today there are many efforts to enable organizing and querying data in Grid environment. Open Grid Forum define a set of specifications which are implemented an in use. OGSA-DAI is a good example.

And, finally, there is a still need for improvements and space for discovering new technologies for working with large amounts of data especially in the field of Grid data mining. We propose several directions for future development.

References

- 1. J. Taylor, Defining e-Science, http://www.nesc.ac.uk/nesc/define.html.
- 2. G. Kakasevski, et al., Grid enabled system for gathering and retrieving images from WEB, ETAI 2007.
- 3. The Large Hadron Collider Homepage. http://lhc.web.cern.ch/lhc/.
- 4. I. Foster, C. Kesselman. The Grid: Blueprint for a New Computing Infrastructure. Elsevier, 2004.
- P. Brezany, I. Janciak, and A.M. Tjoa, "GridMiner: A Fundamental Infrastructure for Building Intelligent Grid Systems," Proc. 2005 IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI'05), 2005, pp. 150--156.
- Vlado Stankovski, Martin T. Swain, Valentin Kravtsov, Thomas Niessen, Dennis Wegener, Jörg Kindermann, Werner Dubitzky: Grid-enabling data mining applications with DataMiningGrid: An architectural perspective. Future Generation Comp. Syst. 24(4): 259-279 (2008).
- Congiusta, D. Talia, and P. Trunfio, "Distributed data mining services leveraging WSRF," Future Generation Computer Systems, vol. 23, no. 1, 2007, pp. 34--41.
- 8. D. Guedes, W.Jr. Meira and R. Ferreira, "Anteater: A Service-Oriented Architecture for High-Performance Data Mining", IEEE Internet Computing, 2006, pp. 36--43.
- Rahul Ramachandran, Sara Graves, John Rushing, Ken Keizer, Manil Maskey, Hong Lin and Helen Conover; ADaM services: Scientific data mining in the service-oriented architecture paradigm; Data Mining Techniques in Grid Computing Environments 2008.
- 10.N.P.C. Hong, et al., Grid Database Service Specification, GGF Informational Document, Feb. 2003.
- 11.Kanishka Bhaduri, Kamalika Das, Kun Liu, Prof. Hillol Kargupta; Distributed Data Mining Bibliography; July 2008.
- 12.I. Foster, C. Kesselman, J. M. Nick, and S. Tuecke, "The physiology of the grid: An open grid services architecture for distributed systems integration," in Open Grid Service Infrastructure WG, Global Grid Forum, vol. 22, pp. 1–5, Edinburgh, 2002.
- Open Grid Forum community of users, developers, and vendors for standardization of Grid computing; http://www.ogf.org/; 2012
- 14.Atkinson, M.P., Dialani, V., Guy, L., Narang, I., Paton, N.W., Pearson, D., Storey, T. and Watson, P. Grid Database Access and Integration: Requirements and Functionalities, DAIS-WG, Global Grid Forum Informational Document (GFD.13), March 2003.
- 15.P. Watson, "Databases and the Grid," in Grid Computing: Making the Global Infrastructure a Reality, F. Berman, G. Fox, and A. J. G. Hey, Eds.: Wiley, 2002, pp. 1060.
- 16.A. Lee, J. Magowan, P. Dantressangle and F. Bannwart. Bridging the Integration Gap, Part 1: Federating Grid Data. IBM Developer Works. August 2005.
- 17.Steven Lynden, Mario Antonioletti, Mike Jackson, Sunil Ahn; WS-DAI and WS-DAIR Implementations Experimental Document; DAIS Working Group 2009.
- Nuno Santos, Birger Koblitz; Distributed Metadata with the AMGA Metadata Catalog; CoRR abs/cs/0604071 2006.
- 19.M. Antonioletti, M. Atkinson, R. Baxter, A. Borley, N. Hong, B. Collins, N. Hardman, A. Hume, A. Knox, M. Jackson, et al., "The design and implementation of Grid database services in OGSA-DAI," Concurrency and Computation: Practice and Experience, vol. 17, no. 2-4, pp. 357–376, 2005.
- 20.Tanimura Y, Yamamoto N, Tanaka Y, Iwao K, Kojima I, Nakamura R, Tsuchida S and Sekiguchi S. Evaluation of Large-Scale Storage Systems for Satellite Data in GEO GRID. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B4. pp 1567-1574. Beijing 2008.
- 21.MESSAGE (Mobile Environmental Sensing System Across Grid Environments) Project. http://bioinf.ncl.ac.uk/message/. 2010.

- 22.Jackson, M. et al. OGSA-DAI 3.0-the whats and the whys. In Proc. UK e-Science All Hands Q6 Meeting pp. 158–165, 2007.
- 23.OGSA-DAI open source project publications. http://sourceforge.net/apps/trac/ogsadai/wiki/Publications. 2012.
- 24.Dobrzelecki B, Krause A, Hume A, Grant A. Antonioletti M, Alemu Y, Atkinson M, Jackson M, Theocharopoulos E (2010) Integrating distributed data sources with OGSA–DAI DQP and Views. Phil. Trans. R. Soc. A 13 September 2010 vol. 368 no. 1926 4133-4145. 2010.
- 25.Lynden, S., Mukherjee, A., Hume, A. C., Fernandes, A. A., Paton, N. W., Sakellariou, R., and Watson, P. 2009. The design and implementation of OGSA-DQP: A service-based distributed query processor. Future Gener. Comput. Syst. 25, 3 (Mar. 2009), 224-236. 2008.
- 26. The Design and Implementation of OGSA-DAI-DM; Yan Zhang, Honghui Li, Gang Dai; CSEEE 2011.
- 27.Anastas Misev, Emanouil I. Atanassov: User Level Grid Quality of Service. LSSC 2009: 507-514.
- 28.A Distributed Architecture for Data Mining and Integration; M.P. Atkinson et al.; 2009.

An Ontology-driven Approach to Self-management in Cloud Application Platforms

Rustem Dautov, Iraklis Paraskakis, Dimitrios Kourtesis

South-East European Research Centre (SEERC), City College – International Faculty of the University of Sheffield, 24 Proxenou Koromila Street, 54622 Thessaloniki, Greece {rdautov, iparaskakis, dkourtesis}@seerc.org

Abstract. Cloud application platforms, with their support for elastic resources and software development lifecycle, are attracting more and more users wishing to host Web-based applications that have variable demand, yet consistent performance requirements. Effective platform and application management is mandatory in order for applications running in the cloud to be cost-effective. We see self-management as one of the ways of achieving this goal. In this paper we present our vision of an ontology-driven approach to self-management in cloud application platforms. In the context of the MAPE-K reference model for adaptation cycles, we describe main advantages of using ontologies and rules, in order to represent self-reflective knowledge and define adaptation policies in comparison to more traditional approaches.

Keywords: Ontology, self-management, cloud application platform, adaptation, self-reflection.

1 Introduction

Cloud application platforms aim significantly to reduce the effort required to create a Web application and deploy it to a cloud, and are completely eliminating the effort needed to maintain a hosting environment [1]. Managing the environment and ensuring that applications operate at a satisfactory level becomes a concern of the platform. In most cases, to ensure that appropriate service levels are maintained, cloud platforms are continuously monitoring the usage of deployed applications and are capable of reacting to important changes in the environment by carrying out some forms of automatic adaptation [1]. Those forms of adaptation, however, only deal with changes at the level of infrastructure services, and are therefore rather elementary.

This paper introduces some benefits of using semantic technologies in the context of self-management in cloud application platforms, and sketches a possible approach to supporting self-management through the realization of an ontology-based adaptation framework. The research work introduced in this paper is novel, because this topic has not yet been addressed in the literature, and we consider that the application of ontologies in the context of self-managing cloud platforms is a research area worth studying.

The paper is structured as follows: section 2 introduces the topic of cloud application platforms and self-management. Section 3 presents ontologies and the benefits of utilising them in the context of self-managing cloud platforms. Section 4 is dedicated to a high-level description of an ontology-driven approach to self-management in cloud platforms. Section 5 describes the related work in brief. Finally, in section 6, we summarize the previous work and outline the future research steps.

2 Background and Motivation

2.1 Cloud Application Platforms

Cloud application platforms (henceforth - cloud platforms) are transforming the processes of development, deployment and maintenance of Web-based software applications [2]. Platforms such as Google App Engine, Microsoft Azure, Force.com, Heroku, and other contemporary Platform-as-a-Service (PaaS) offerings aim at supporting the complete lifecycle of building and delivering Web applications and facilitate the deployment of software applications without the necessity of buying and managing the underlying hardware and software layers [2, 3].

Various PaaS vendor offerings differ with respect to the integrated facilities they provide to users. Some of them are just "bare" platforms with a very basic set of services. For example, Amazon Elastic Beanstalk [4] provides software developers only with an operating system, a Web container for PHP or Java applications, and some basic configuration options. It is the developers' responsibility to develop and to test an application remotely, deploy it to the cloud service, and maintain it during runtime. On the other hand, there are platforms supporting the complete software development life-cycle and even liberating their customers from routine tasks of developing, testing, deploying, etc. For example, with Zoho [5], users do not even need to code, as configuration and deployment of an application can be performed through a "point-and-click interface". Other PaaS offerings can provide their users with facilities and on-demand tools enabling several or all of the following features [3]:

- Application design, development, testing and deployment
- Database, storage, and persistence integration
- Bug tracking
- Team collaboration
- Application versioning
- Web service integration
- Compatibility with other cloud platforms
- State management
- Developer community support
- Auto-scaling, etc.

Regardless of the particular functionalities being offered, a common characteristic of cloud platforms is that they reduce the effort required to create a Web application, to deploy it to a hosting environment, and to maintain an underlying infrastructure [6]. Monitoring the infrastructure and ensuring that deployed applications fully function at a satisfactory level becomes a concern of the platform, rather than of developers and administrators [1]. To provide that functionality, advanced cloud platforms are equipped with mechanisms responsible for carrying out different forms of adaptation.

2.2 Cloud Platform Adaptation Mechanisms

To ensure that appropriate service levels are maintained, cloud platforms are continuously monitoring the usage of deployed applications, and are capable of reacting to important changes in the environment by carrying out some forms of automatic adaptation. Most commonly, cloud platforms are capable of adapting to varying volumes and types of user requests by allocating the incoming workload across different computational instances (load balancing), or by reserving and releasing computational resources upon demand (elasticity) [2, 3]. Load balancing and elasticity are both essential properties of a cloud service, according to the National Institute of Standards and Technology (NIST) [7].

Monitoring various system characteristics during the execution of applications (e.g., response time, throughput, utilization of virtual machines and containers, etc.), and generating actions accordingly is a common mechanism in existing PaaS offerings. A mechanism of this form may provision or remove virtual machines and program licenses, allocate containers within virtual machines, launch additional threads for faster execution [8], etc.

Together with adaptive actions performed at the Infrastructure-as-a-Service (IaaS) level (e.g., provisioning and removing computational instances, network management, etc.), such a framework mainly deals with scalability and seems to be the only adaptive mechanism employed by most cloud platforms [3].

2.3 Limitations of the State of the Art in Adaptation Mechanisms

It becomes apparent that the forms of adaptation in existing cloud platforms are only dealing with changes at the level of infrastructure services, and are therefore rather elementary. More complex and intelligent adaptation scenarios, such as the need to modify the actual structure and/or behavior of an application during its runtime, are much more difficult to automate, and are presently beyond the capabilities of common cloud platforms.

At the age of the Internet of Services, when applications are more and more depending on third-party services [9], a failure of an external component at one point may lead to malfunctioning of a whole cloud-based system, without the platform noticing it. Presently, cloud platforms are incapable of detecting and reacting to such situations. Cloud platforms cannot make a distinction between infinitely looping and long-running operations, cannot understand whether a response from an external service is correct or not, cannot substitute a Web service with another, etc. These limitations make it necessary for platform administrators and application developers to be involved in the lifecycle of an application even after it has been deployed to a

cloud environment. That is, the adaptation is done manually by rewriting the application/platform code, recompiling, redeploying the application or restarting the platform.

All these aforesaid limitations bring us to an idea of developing a *self-managing* cloud platform - what if we could enhance a cloud platform with an *autonomic* adaptation framework?

2.4 Self-management and Self-reflective Systems

According to IBM's view on autonomic computing, the following fundamental features are usually considered under the umbrella term of *self-management* [10, 11]:

- *Self-configuration*: the ability of a system to adapt automatically to dynamically changing environment
- Self-healing: the ability of a system to detect, diagnose and react to failures.
- *Self-optimization*: the ability of a system to efficiently maximize utilization of available resources automatically
- *Self-protection*: the ability of a system to discover and protect themselves from various malicious attacks.

One of the possible ways of achieving self-management in a cloud platform is through self-reflection. [12]. A self-reflective system refers to the use of a causally connected self-representation to support the inspection and adaptation of that system [12]. It means that such a system is self-aware of its internal structure (i.e. system subcomponents, interconnections between subcomponents, available resources, etc.) and able to perform run-time adaptations, so that applied adaptations dynamically reflect on the state of the system (thus, possibly, triggering another adaptation cycle). The motivation behind self-reflection stems from the necessity to have systems that, when deployed in hostile and/or dynamically changing settings, are capable of reacting to various changes in the environment. In such scenarios, the capability of a remote system to perform automatic adaptations at run-time, and within a specific time frame, is often of a great importance.

3 Ontologies as a Way of Achieving Self-reflection

3.1 Ontologies

Ontologies, in their contemporary sense, started attracting attention in the 1970s as a way of representing and reasoning about knowledge in the context of artificial intelligence. An ontology, as it is widely used in the modern computer science literature, is "a formal, explicit specification of a shared conceptualization" [13]. Conceptualization refers to an abstract model of a phenomenon existing in the real world. This abstract model includes only relevant concepts of that phenomenon. Explicit means that the type of concepts used, and the constraints on their use are explicitly defined. Formal means that an ontology should be unambiguous and

machine-readable. *Shared* refers to the fact that knowledge comprising an ontology is accepted and agreed on by a group of people, not just an individual [13].

By means of ontology specification languages, such as the Web Ontology Language (OWL) [14], it is possible to explicitly define knowledge (i.e. concepts, relations, properties, instances, etc.) and basic rules in order to reason about this knowledge. Rule specification languages, such as the Semantic Web Rule Language (SWRL) [14], extend ontology specification languages by providing a way of defining more complex and expressive rules. Based on explicitly defined knowledge and rules, a reasoning engine can infer additional, implicit knowledge.

Fields where ontologies are widely and successfully used are enterprise interoperability [14, 15], life sciences such as biomedicine [14], the Semantic Web [16], e-commerce [14], and more.

3.2 Potential Benefits of Using Ontologies to Support Self-reflection

Utilising ontologies to support self-reflection may be one of the possible solutions to the above-mentioned problem of ineffective adaptation at the PaaS level. An ontology-driven adaptation framework would allow platform administrators to create high-level specifications of a platform's internal structure as well its adaptation policies by means of ontologies and rules [17]. Given a set of such policies, an adaptation engine would reason, at runtime, about whether or not some adaptation needs to be performed (e.g. substitution of an external Web service) in response to particular usage conditions (e.g. the user's context, the combination of applications being used, the types of data involved, etc.).

Compared to more traditional methods explored in the autonomic computing literature [10, 11], an ontology-based approach to platform self-management could provide the following benefits for application developers and platform administrators:

- Separation of concerns: with ontologies separated from the platform/application programming code, it is easier to make changes to adaptation policies on the fly (i.e. without recompiling, redeploying and restarting the platform/application) and to maintain the adaptation framework. Moreover, this approach allows non-professional programmers (i.e. domain specialists) to develop adaptation policies as well.
- More flexible and intelligent adaptation at any level of abstraction: it is up to platform designers to decide whether adaptation actions will be carried out at the lowest level of programming classes and variables or at the upper-most level of the whole platform. It is worth mentioning here that the ontology-based approach is generic and could be applied to adaptations at the IaaS level as well. Moreover, this approach could be potentially applied not only to cloud platforms, but to other distributed systems (e.g. service-based applications).
- Increase in reuse, automation and reliability: once implemented and published on the Web, ontologies are ready to be re-used by any third parties, thus saving ontology engineers from "reinventing the wheel". Since the reasoning process is automated and performed by a reasoning engine, it is not prone to so-called "human factors" and more reliable (provided the correctness and validity of ontologies and rules).

4 Towards self-reflection in cloud-platforms

4.1 The MAPE-K Autonomic Loop

Taking into consideration all aforementioned issues, we propose to investigate an ontology-driven approach to achieve self-management in cloud application platforms. This approach implements IBM's reference model for autonomic control loops [18], known as the MAPE-K (Monitor, Analyse, Plan, Execute, Knowledge) loop. Figure 1 depicts this generic reference model.

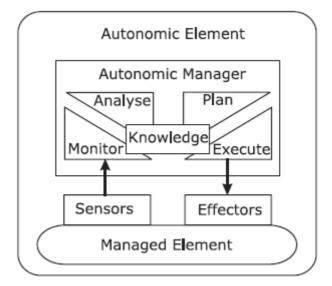


Fig. 1. IBM's MAPE-K reference model for autonomic control loops (excerpted from [18]).

Let us examine this model in the context of adaptations in a cloud platform. The *autonomic element* represents a cloud platform, whereas the *managed element* could be either one of the deployed applications or the platform itself. The managed elements are equipped with *sensors*, responsible for collecting information them, and *effectors*, responsible for executing instructions to adapt them.

The *autonomic manager* is a software component and the core element of the adaptation mechanism. Its main responsibilities are:

- (i) *monitoring* the managed elements by means of collecting data from the sensors;
- (ii) *analysing* the data collected from the sensors;
- (iii) *planning* required adaptation activities based on the analysis;
- (iv) *executing* the adaptation plan.

All these activities are done based on the internal (self-reflective) *knowledge* of the system. The knowledge may also include a set of policies expressed in the form of *"event-condition-action"* (ECA) rules [18]. That is, a typical policy could be represented as a following rule: "When an *event* occurs and a *condition* holds, then an *action* is executed".

Let us consider a scenario of an application system consuming a potentially malfunctioning Web service. In this case, the managed element is the application system, and the result returned by the service is one of the parameters being monitored. The autonomic manager possesses knowledge about the structure of the application and its connection to the Web service, as well as the list of equivalent services that can act as substitutes. The knowledge base, among other things, also includes a policy stating that if the service returns an incorrect result, then this service needs to be replaced.

4.2 Role of Ontologies

So how could ontologies fit into this MAPE-K model? Given the characteristics of ontologies, as explained briefly in section 3.2, we believe that it is worthwhile investigating the benefits (and drawbacks) of using ontologies for representing internal knowledge of a managed system and reasoning over this knowledge.

As the main source of self-reflective knowledge comprising the knowledge base of the autonomic manager, ontologies could provide formal models of the internal structure of the application platform (e.g. its components and relations between them, concepts related to the adaptation process, etc.) and define policies for carrying out adaptation activities. This is where a major benefit of ontology-driven approach becomes obvious. We, as platform developers, are exempted from the time-consuming and effort-intensive task of implementing our own (potentially error-prone) analysis engine from scratch, because an ontology-driven approach already provides us with this functionality by means of reasoning engines. It means that a routine for reasoning over (i.e. analysing) a set of situations and adaptation alternatives, is done by an already existing, resilient, and efficient mechanism. In contrast to this approach, consider a scenario where the self-reflective knowledge would be stored in a database or XML files. In this case, we would need to "hard-code" the adaptation logic with numerous "*if-then*" and "*switch-case*" operators.

We now consider some simple examples illustrating the self-managing capabilities of an ontology-driven approach. Please note that these examples only serve the purpose of demonstrating the idea of using ontologies and rules for self-adaptation, and are therefore very simple. In practice, the set of ontologies and rules could be much more complex.

Let us consider an OWL ontology that defines a vocabulary of concepts related to the system and the adaptation process. Among other things, it includes the classes *Service*, *Time*, *System*, and *Client*. The terms *hasResponseTime*, *hasHighResponseTime* and *hasCrashed* correspond to data properties in the ontology, and *isEquivalent* and *needsSubstitution* are object properties whose domain is the class *Service*. The terms *isRequestedBy* is an object property of the class *System*, and *requestsPerSecond* and *isMalicious* are data properties of the class *Client*. Let us start with a policy stating that if a response time from a Web service is greater than 5 seconds, that service can be considered to have crashed and should be substituted. This policy, represented in SWRL, could look like (1). Following these rules, the autonomic manager can deduce that a service should be substituted, thus enabling the platform to demonstrate self-healing capabilities.

```
Service(?s1) ∧ Time(?t) ∧ hasResponseTime(?s1, ?t) ∧
greaterThan(?t, 5000) → hasCrashed(?s1, true)
hasCrashed(?s1, true) ∧ Service(?s2) ∧ isEquivalent(?s1, ?s2)
→ needsSubstitution(?s1, ?s2)
```

The following rules (2) imply that if a Web service does not reply in 2 seconds, it is considered to have a high response time and might need to be substituted. Based on this rule, the adaptation framework may optimize the system by replacing a "slow" service with a "faster" one (if there is any).

```
Service(?s1)
                 \wedge
                       Time(?t)
                                    \wedge
                                        hasResponseTime(?s1,
                                                                     ?t)
                                                                            \wedge
greaterThan(?t,
                       2000)
                                           lessThan(?t,
                                                               5000)
                                                                           \rightarrow
                                   \wedge
hasHighResponseTime(?s1, true)
                                                                                 (2)
hasHighResponseTime(?s1,
                                   true)
                                               \wedge
                                                       Service(?s2)
                                                                            \wedge
hasHighResponseTime(?s2, false) ^ isEquivalent(?s1, ?s2)
                                                                           \rightarrow
needsSubstitution(?s1, ?s2)
```

As an example of self-protecting capabilities, we can consider a rule defining a Denial-of-Service (DoS) attack as a case when the number of incoming external requests per second reaches 1000 (3), and these requests come from the same address. Given such a rule, the adaptation framework is able to identify potential threats and block the malicious IP address.

```
System(?s) \land Client(?c) \land isRequestedBy(?s, ?c) \land (3) requestsPerSecond(?c, 1000) \rightarrow isMalicious(?c, true)
```

As illustrated by these examples, the ontology-driven approach implies that ontologies should be used not only to define a common vocabulary, but also to provide the "building blocks" for constructing adaptation policies (i.e. concepts, relations, properties, etc.). It also becomes obvious that with the programming code and the policies separated from each other, we can benefit from much easier maintenance and evolution of the policies, because any changes to ontologies and rules will not affect the rest of the system.

4.3 Potential Challenges and Research Questions

Our research work is still at the initial stage, and presently it is hard to foresee all the challenges that will arise. Nevertheless, some important issues to be studied are already evident.

A delay between the moment when the system identifies a problem and the moment when an adaptation actually takes place, is a common problem in real-world systems [18]. In frequently changing environments, adaptation actions may become irrelevant by the time they are applied to the system, and this is an important challenge to be addressed. This problem has already been studied in the research field of Complex Event Processing and Stream Processing – processing of large, continuously flowing data streams [19]. A related research topic, Stream Reasoning, goes one step further and combines techniques from stream processing with ontology engineering in order to support "on the fly" reasoning over incoming streams of semantically-enriched data [20]. Stream reasoning deals with events - something that happens or changes in the current state of affairs [20]. In the context of selfmanagement, an event can be a failure of a component, a response from a Web service or even a daylight saving time shift. In order to be processed, these events need to be somehow converted into computational models. However, in our case we also wish to reason about these events, and that is why ontologies could be used for representing events in stream reasoning.

Another challenge, also emerging from the area of stream reasoning, is identifying the parameters to be monitored. That is, which parameters are required to provide effective adaptation, and which parameters may be neglected? In a scenario where the managed frequently-changing system supplies the autonomic manager with large quantities of data, it is essential to distinguish between the valuable information and so called "noise" [19].

This brings us to another research challenge - the minimization of the impact of the autonomic manager's presence on the overall performance of the platform. In other words, with many applications running and being monitored, the autonomic manager may become the platform's "bottleneck". We have to propose replication mechanisms (i.e. launching additional instances of the autonomic manager) and load balancing strategies for spreading the load across several instances of the autonomic manager. At this stage of the research, it is not yet clear if this will be the case, but this may be an important aspect to investigate.

Another research question concerning the future implementation of the autonomic manager is how it will interact with the platform and deployed applications within the cloud platform. That is, how the sensors and effectors will be designed and implemented. Presumably, we will need to propose an API that developers could follow when designing and implementing applications. This API would provide some entry-points for the autonomic manager to perform so called *active* monitoring [18].

5 Related Work

The research topic described so far is positioned at the intersection of three major research areas: Cloud Computing, Autonomic Computing, and Ontology Engineering. A literature survey in this direction has shown that our work seems to be novel, and apparently there has not been any research done in the direction of using ontologies in the context of autonomic cloud platforms.

Nevertheless, there have been research activities that cover two of the abovementioned research areas and provide potentially valuable insights. For instance, much effort has been expended in order to develop autonomic clouds, although most researchers focus on achieving self-management at the IaaS level (e.g. [21–23]). Especially valuable and relevant research work was done by Maurer et al. [23], who also adapted the MAPE-K model in order to achieve autonomic management of cloud infrastructures and prevent SLA violations. In the context of the Foundations of Self-Governing ICT Infrastructures (FoSII) project, they have adopted Case Based Reasoning (CBR) and Rules-based Systems in order to support knowledge management at the PaaS level and presented a vision of an autonomic self-aware cloud market platform. Brescovic et al. also employ the MAPE-K reference model in their work. They introduce an autonomic cloud market monitoring methodology and demonstrate its viability with relevant use-cases.

There have also been some research activities that cover the application of ontologies to autonomic computer systems. For example, Carey et al. [25] were among the first to employ ontologies to define adaptation policies, to support autonomic communications in ubiquitous computing environments. A similar approach has been described by Vassev and Hinchey [26]. In their survey the authors provide a comprehensive analysis of different approaches to self-reflective knowledge representation, and describe their experience of utilising ontologies to capture the knowledge domain and adaptation policies in the context of the Autonomic Service-Component ENSembles (ASCENS) project. Söldner, Kapitza, and Meier [27] argue that existing approaches to adaptive software lack a semantic model to define and perform adaptations in a business-oriented way, and introduce programming support for context-aware adaptations that is built upon an OWL-based semantic model. Although [25], [26] and [27] are mainly focused on research fields distant from cloud computing, the described experience and results of utilising ontologies to achieve self-reflection seem to be promising.

6 Conclusions and Future Work

In this paper we gave a high-level description of an ontology-driven approach to creating a self-managing cloud platform. The approach builds on IBM's MAPE-K reference model, which is a widely used generic model for designing closed adaptation loops. Having introduced some theoretical background on the topics of autonomic computing and ontology engineering, we introduced an idea of how ontologies and rules could be utilised in order to represent the internal knowledge of the platform and provide reasoning over this knowledge, thus exempting platform developers from implementing their own analysis engine. Other benefits of using ontologies, such as the separation of concerns, increased reuse, automation and reliability, were described as well.

The future work includes a deeper study of theoretical underpinnings relevant to this project, as well as implementation details of existing adaptation engines. Particularly promising directions of the further investigation are Stream Reasoning and Complex

Event Processing (CEP), as these topics may provide several lessons to be learnt in the context of research challenges. This (yet, not exhaustive) list includes the delay in adaptation actions, the selection of parameters to be monitored, the way the autonomic manager will interact with a cloud platform, and minimizing the impact of the manager's presence in the system. A survey of existing adaptation mechanisms from a perspective of their applicability to cloud platform-based scenarios may be another potential source of insights and valuable research experience.

Acknowledgments. This research work is funded by the FP7 Marie Curie Initial Training Network RELATE on Engineering and Provisioning Service-based Cloud Applications.

References

- 1. Shroff, G.: Enterprise Cloud Computing: Technology, Architecture, Applications. Cambridge University Press (2010).
- Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M.: Above the Clouds: A Berkeley View of Cloud Computing, (2009).
- Natis, Y.V., Knipp, E., Valdes, R., Cearley, D.W., Sholler, D.: Who's Who in Application Platforms for Cloud Computing: The Cloud Specialists. Gartner Research (2009).
- 4. Amazon Elastic Beanstalk, http://aws.amazon.com/elasticbeanstalk/.
- 5. Zoho, http://www.zoho.com/.
- Dillon, T., Chen Wu, Chang, E.: Cloud Computing: Issues and Challenges. 2010 24th IEEE International Conference on Advanced Information Networking and Applications (AINA). pp. 27–33. IEEE (2010).
- 7. Mell, P., Grance, T.: The NIST definition of cloud computing. National Institute of Standards and Technology. 53, 50 (2009).
- Litoiu, M., Woodside, M., Wong, J., Ng, J., Iszlai, G.: A business driven cloud optimization architecture. Proceedings of the 2010 ACM Symposium on Applied Computing. pp. 380–385. ACM, New York, NY, USA (2010).
- 9. Erl, T.: Service-Oriented Architecture: Concepts, Technology, and Design. Prentice Hall PTR, Upper Saddle River, NJ, USA (2005).
- Ganek, A.G., Corbi, T.A.: The dawning of the autonomic computing era. IBM Systems Journal. 42, 5–18 (2003).
- 11. Kephart, J.O., Chess, D.M.: The vision of autonomic computing. Computer. 36, 41–50 (2003).
- 12. Blair, G.S., Coulson, G., Grace, P.: Research directions in reflective middleware: the Lancaster experience. Proceedings of the 3rd workshop on Adaptive and reflective middleware. pp. 262–267. ACM, New York, NY, USA (2004).
- Studer, R., Benjamins, V.R., Fensel, D.: Knowledge engineering: Principles and methods. Data & Knowledge Engineering. 25, 161–197 (1998).

- 14. Hitzler, P., Krötzsch, M., Rudolph, S.: Foundations of Semantic Web Technologies. CRC Press (2009).
- 15. Uschold, M., Gruninger, M.: Ontologies and semantics for seamless connectivity. SIGMOD Rec. 33, 58–64 (2004).
- 16. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American. (2001).
- 17. Stojanovic, L., Schneider, J., Maedche, A., Libischer, S., Studer, R., Lumpp, T., Abecker, A., Breiter, G., Dinger, J.: The role of ontologies in autonomic computing systems. IBM Systems Journal. 43, 598–616 (2004).
- 18. Huebscher, M.C., McCann, J.A.: A survey of autonomic computing—degrees, models, and applications. ACM Comput. Surv. 40, 1–28 (2008).
- 19. Barbieri, D., Braga, D., Ceri, S., Della Valle, E., Grossniklaus, M.: Stream Reasoning: Where We Got So Far. Proceedings of the 4th International Workshop on New Forms of Reasoning for the Semantic Web: Scalable and Dynamic (NeFoRS) (2010).
- Anicic, D., Fodor, P., Rudolph, S., Stojanovic, N.: EP-SPARQL: a unified language for event processing and stream reasoning. Proceedings of the 20th international conference on World wide web. pp. 635–644. ACM, New York, NY, USA (2011).
- Martin, P., Brown, A., Powley, W., Vazquez-Poletti, J.L.: Autonomic management of elastic services in the cloud. Computers and Communications (ISCC), 2011 IEEE Symposium on. pp. 135 –140 (2011).
- 22. Casalicchio, E., Silvestri, L.: Architectures for autonomic service management in cloud-based systems. Computers and Communications (ISCC), 2011 IEEE Symposium on. pp. 161–166 (2011).
- 23. Maurer, M., Breskovic, I., Emeakaroha, V.C., Brandic, I.: Revealing the MAPE loop for the autonomic management of Cloud infrastructures. Computers and Communications (ISCC), 2011 IEEE Symposium on. pp. 147–152 (2011).
- Breskovic, I., Haas, C., Caton, S., Brandic, I.: Towards Self-Awareness in Cloud Markets: A Monitoring Methodology. Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on. pp. 81–88 (2011).
- Carey, K., Courtenage, S., Feeney, K., Lewis, D., Tiropanis, T.: Semantic-Based Policy Engineering for Autonomic Systems. IFIP Lecture Notes in Computer Science (LNCS). 3457, 152–164 (2005).
- Vassev, E., Hinchey, M.: Knowledge Representation and Awareness in Autonomic Service-Component Ensembles-State of the Art. Object/Component/Service-Oriented Real-Time Distributed Computing Workshops (ISORCW), 2011 14th IEEE International Symposium on. pp. 110– 119 (2011).
- 27. Söldner, G., Kapitza, R., Meier, R.: Providing Context-Aware Adaptations Based on a Semantic Model. In: Felber, P. and Rouvoy, R. (eds.) Distributed Applications and Interoperable Systems. pp. 57–70. Springer Berlin / Heidelberg (2011).

Cloud Computing and E-Learning: a Review

Krenare Pireva¹, Petros Kefalas² and Tony Cowling³

 ¹ South-East European Research Centre, Thessaloniki, Greece krpireva@seerc.org
 ² The University of Sheffield International Faculty, CITY College, Greece

kefalas@city.academic.gr

³ The University of Sheffield, UK

A.Cowling@dcs.shef.ac.uk

Abstract. With significant recent advances in information and communication technology, many education institutions are reconsidering their computing infrastructure requirements. The cloud seems to offer a cheaper and more reliable alternative for purchasing and maintaining expensive hardware and developing new local software services. Cloud computing is expected to be a smart alternative to start-up and maintain educational establishments since the cloud infrastructure is robust and always available. E-learning is one of the areas that could benefit from cloud computing. The aim of this paper is to outline the reasons behind this technology transfer as well as discuss additional potential benefits and setup future research paths.

Keywords: e-learning, cloud computing, services, education

1 Introduction

E-learning facilitates the educational process for students who are not able to attend face to face lectures in a variety of ways, such as:

- access to lecture materials;
- synchronous and asynchronous communication with lecturers;
- communication and interaction with fellow students;
- access to libraries and variety of other media;
- access to other educational services and learning technologies etc.

People's educational needs differ; for instance, students have different learning styles[1] while teachers apply different teaching methodologies. E-learning, as opposed to traditional learning, offers the opportunity to personalize their learning environment, which can be accessed anytime, anywhere. Availability of a variety of on-line material contributes to the enhancement of quality of learning. In addition, as learning can also be seen as the product of interaction [2], e-learning applications (in the form of networked learning) maximize the value of collaboration and interaction which in turn increases the learning opportunities. New learning technologies allow for collaboration on real-life situations and usually motivate co-operation between groups of people who develop a new learning culture [3]. E-learning platforms, such as virtual learning environments (VLE) [4] are concerned with enhancing students capabilities with collaboration and interaction by offering a variety of techniques that facilitate the learning process [5]. In virtual learning environments students are actors instead of being only active [6], they support social interaction in many ways. Many institutions have adopted VLEs to complement traditional teaching, and such action is well received by staff and students [7].

Setting up a VLE comes with the cost for buying new hardware and software for a specified number of users, engaging support people for administering and continually managing the resources based on students activities. In order to provide satisfactory e-learning services, an institution needs to setup the VLE based on peak demand of the system. However, by doing this, the system is unevenly utilized since there are only few peaks of learning resources demand in different periods of the year. This necessarily leads to extra investment, since most of the VLEs do not support dynamic allocation for computing and storage capacity. They fail to engage students with e-learning activities and sometimes this leads to data duplication [8]. Therefore, it is a natural consequence that today e-learning systems can use the benefits of cloud computing technology in all service models [9, 10]. However, apart from the obvious, there might be more benefits from getting e-learning into the cloud. The cloud, apart from storage, scalability, accessibility and reliability, offers the opportunity to have sharing among a wider group of people as well as a plethora of new services.

This paper will discuss the impact of cloud computing on e-learning. We will start by reviewing public and private cloud computing, different cloud service models and current services that have been deployed [11-14]. Additionally, we will review the impact of e-learning in education systems with emphasis on how e-learning increases the collaboration of students and how its impact changes the learning process compared to traditional learning. Further to the above, we will make an initial investigation of what additional benefits e-learning could offer through such transition. We will argue whether there are any impacts on increasing the interactivity and collaboration of the e-learning systems and how this approach will enable access to resources, services, curricula and a variety of other features which could enhance the learning experience. For example, it may be possible that cloud could widen the participation of educators and staff by forming a learning community which is not confined within a specific institution. Finally, we plan to define a few directions of research for an innovative framework, namely "Cloud e-learning". Based on cloud characteristics and benefits, it will aim to deliver more reliable, scalable and cost effective as well as innovative learning services to educators and students.

2 Cloud Computing

Cloud Computing has been coined as the fifth generation of computing, leaving behind: mainframes, personal computers, client servers and web services [15].

It offers IT capabilities (hardware, software, services) to different institutions which want a cost-effective and efficient cloud platform.

Different computing researchers have tried to define cloud computing in their own way [16], however the main idea of cloud computing is to deliver IT tasks "as a service" through its resource pool. It contains different network devices for providing services to the users. Cloud computing could be called elastic since the users can start, stop and create virtual machines through web services [17–19]. The main idea behind the cloud is how the Internet is presented in different network diagrams. It could be considered as an access point which has a lot of different network devices, which respond instantly to all client requests. The main idea of cloud computing is that cloud offers services for renting and the clients are the customers which can pay per their services.



Fig. 1. An Example of cloud computing infrastructure

Report [20] defines cloud computing as "a style of computing where massively scalable IT-related capabilities are provided "as a service" using Internet technologies to multiple customers". Figure 1 depicts a cloud computing example, which illustrates the idea that a cloud contains different network devices, where clients can access the services through PC, smart phones, PDAs etc. All clients are classified into four categories [21]

- Zero clients which are defined as devices with no operating system, they connect to the network through power-on;
- Mobile clients which includes all smart phones or PDAs;
- Thin clients which includes computers that do not have internal hard drivers, but rather let the server to do the work, and then displays the information; and
- Thick clients which includes the usual computers which uses any web browser to connect to the cloud.

For e-learning, the third category, thin clients, are becoming very popular because of lower hardware cost, lower IT cost, data security, less power consumption etc. Quality of service, stretch of the resources and the movement of the processing efforts from the local machines to the data center systems are some of many key factors why cloud computing is becoming so attractive for e-solutions, taking into account the e-learning solutions. Other factors which could be counted as advantage comparing to the non cloud computing technology are: the reliability, scalability, portability and low cost. While, as disadvantage could be counted the speed of Internet connection which may affect the connectivity performance, security of the data, the lack of standard protocols which defines communication and inter-work between clouds and cloud computing layers.

The major technology companies are becoming also the first cloud computing players. Among them are, Google, Amazon, Microsoft, Yahoo, Intel, IBM, Rackspace, salesforce.com etc. They already invested in their infrastructure services and now they are providing different service solutions where millions of users are connecting to. Google App is developed based on the model software as a service which covers: messaging, collaboration and security. In messaging we can mention Google talk, Gmail, calendar whereas Google docs, videos and sites provides collaboration. And security covers the encryption part of communication, email security etc Amazon extended amazon web services to Amazon Elastic Compute Cloud which offers services to be rented by the users, to run their own application while using the processing power of amazon. Microsoft has windows Azure which provides a windows platform. Table 1, shows some of the services that are offered from such leading companies. Rackspace [22] is offering hosting and storage solution, IBM [23] is offering hosting and storage in different type of clouds, Google is offering different Google Apps, Google AppEngine [24] and Google docs [25]. Amazon [11], Microsoft [12] and salesforce.com are offering hosting, storage, platform, app services and tools.

Company	Product	Hosting	Storage	Platform	App Services	Tools
Rackspace	Mosso					
IBM	Public and Private Cloud	\checkmark				
Google	Google App Engine	\checkmark	\checkmark	\checkmark	\checkmark	
Amazon	Amazon Web Services	\checkmark	\checkmark		\checkmark	
Microsoft	Azure Services	\checkmark	\checkmark		\checkmark	\checkmark
Salesforce.com	Force.com	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark

 Table 1. Cloud Service Providers

2.1 Cloud Architecture

Cloud architectures play an important role in cloud computing performance while offering their infrastructure services. They address key difficulties surrounding a large number of data processing, such as automatically scalping machine resources based on users demands, finding as many machines as an application needs, coordinating large-scale jobs in different machine resources, switching data process between different machines in case of any machine failures, or even releasing the resources when the jobs are done. In the future, cloud architecture must achieve to decouple the physical resources that are located within the machine as they are working in different physical machines [26].

In cloud computing there are no standards for following one particular cloud architecture model, but we have different architecture approaches depending on the capabilities and services that are going to be delivered. Cloud architecture is divided in two different segments: front-end and back-end. The front-end includes client computers or mobile sophisticated devices which contains all necessary applications for accessing the cloud computing systems whereas back-end covers various cloud computing servers, which could be provided as virtual machines with powerful processing and a huge capacity of storage within server farm.

This makes cloud computing much more sophisticated than traditional models where the server side contained only one powerful server (machine). In [27], the analyzes of cloud computing have been undertaken from the perspective of service-providers.

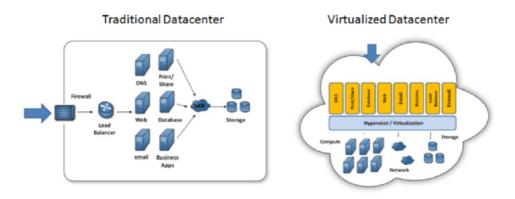


Fig. 2. Traditional versus Cloud datacenter [27]

Different architectural explanations are given in [28], based on user requirements and proposing a comprehensive framework. While choosing an appropriate architecture for the specific e-learning solutions, we should analyze all user-server requirements so the compatibility, availability, compatibility, maintainability, integrity could increase the performance of our services. Viewing from the perspective of user requirements of cloud computing architecture, they would appear to require integrity of data and service integrity.

2.2 Cloud Deployment Models

In cloud computing scenarios, users can distinguish different kinds of cloud. Today there are mainly four different models that cloud providers are deploying their services.

Public Cloud is based on the cloud computing model and usually offers all services to general public. Such cloud are typically owned by large companies, who are able to sell also cloud services over Internet. In the public cloud a key concern is also data security, because the data could be damaged if it is not encrypted while traveling in different locations. These services may be free or they may be charged based on pay-for-use model, on monthly usage per bandwidth and storage. Today, this type of cloud computing has: Amazon [11], Microsoft Azure [12] and Google apps [24] etc.

Private cloud is the most popular cloud for enterprises since the provider has more control on data security, increasing or decreasing the IT capabilities depending on users demand. Usually private clouds are managed by the organization and the data center are on premises [29] which are not becoming available for the public. Organizations access their isolated computing resources via virtual private networks -VPN for extending the existing IT infrastructures.

Community Cloud offers shared infrastructure services by several institutions with specific common concerns, such as: mission, specific security requirements, policy etc. The infrastructure could be located on premises or off premises and they could be managed by the owners or third party.

In **Hybrid Cloud** the infrastructure is a combination of private, community and public clouds that remain unique entities but are bounded by standardized technologies that enable data and application portability.

2.3 Cloud Service Models

Cloud computing is divided in three distinct logical layers [30] based on their main functions: infrastructure, platform and software. Figure 3 gives further explanation of what clients can control or manage and what they cannot. Comparing all three services models, PaaS provides the balance manageability for customers and providers [31, 32].

Infrastructure layer could be viewed as the very bottom logical layer which virtualizes the hardware layer. In this level the system can be scaled based on user requirements. If the system requires more computing power, the system will increase the system resources and vice versa. The infrastructure could be offered as a service and is called as IaaS. There are different virtualization aspects, like virtualizing the operating system, CPU, embedded systems, memory, networks, storage. IaaS cloud customers are able to manage different software systems and different software applications.

Platform as a Service (PaaS) provides resources via a platform where applications can be hosted and developed using programming languages and tools that are supported from the provider. PaaS supplies all resources required to build applications and services from the Internet without having to download

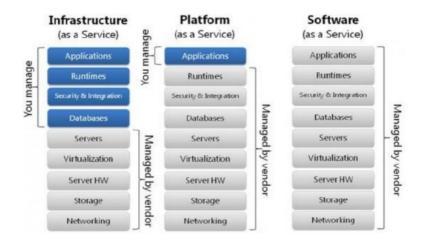


Fig. 3. Cloud Computing Service Models [50]

or install software. Clients that use Cloud PaaS are not able to manage cloud infrastructure, such as: servers, storage, networking, operating system except deployed applications. PaaS services include design development, web service integration, testing, database integration, security, state management, versioning etc. The major PaaS providers are: Force.com, Google App Engine, Microsoft Azure etc. If a specific Platform developer decides to switch the infrastructure vendor than the developer should redesign the part of the applications that rely on the core functionality of the previous vendor [33].

Software as a Service (SaaS) offers the capability to use the provider's application running on the cloud infrastructure. At SaaS level, clients are able to run a provider application, those applications usually are accessed through thin clients interface such as web browser, e.g. Google Mail etc. SaaS services are offered by: Google Docs, Google Calendar etc. As is shown in figure 3, in SaaS model, clients do not have control of Servers, Storage, Networks, Security, platform of applications etc.

3 Cloud Computing vs Conventional Systems

In non cloud computing technologies, institutions owned their hardware, controlled the system, maintained and provided computing capabilities to all levels. For example, within education institutions instead of providing their students with emails they run a server that is owned and maintained by the IT staff employers of institutions and buy software licensed for providing legal services to the students. Difference of cloud computing is that the institutions do not have to own any hardware devices, or to take care of system maintains, because all required hardware and network infrastructure in cloud computing are owned by the third party. So, the education institution through cloud computing can provide emails to students, without having to worry for any software license, buying servers and hiring employees for maintaining infrastructure and the total system.

However, using cloud computing technologies for providing services to end users has some risks that we should note, being aware of possible failures.

- Reliability The Internet connectivity should be sustainable, and in case of any node failure it should be replaced instantly without being noticed from the client part.
- Interoperability Communicating with cloud computing application or platform should not be specific to only one supplier. It is important to note that for current applications and systems it is very limited, and trying to overcome these problems is still a research issue.
- Security The cloud should keep redundant data, so in case of any falsification or theft of information the data should be rolled back by few clicks.

In general while migrating to cloud computing, companies should prepare for most possible failures so the system will have less possibility to fail since it can be prepared to face the major problems which usually cause the failures. Some of characteristics of the cloud computing which favor cloud computing from traditional computing are scalability and flexibility of network, it is location independent, usage measurement services, pay as you use abilities and the most important it is always "ON", which can be accessed anytime from everywhere.

3.1 Technical Benefits of Cloud Computing

Compared with almost any other area, e-learning applications "on the cloud" these days are demanding. They require good performance, high reliability and quality of service (QoS). E-learning systems can use the benefits of cloud computing in all service models. Starting from the infrastructure model, any e-learning system could be provided from the main provider's infrastructure, secondly in the platform as a service model, education centers could use any solution based on the provider's development interface or thirdly the education institutions can decide to use the e-learning as software as a service, which can be accessed via different web clients. Offering e-learning solutions while using cloud computing benefits we can increase the system performance metrics while keeping the e-learning specifications in the top of their performance.

Looking from one perspective, there is no opportunity to determine the place of the servers where the data are being stored and from which server are served, secondly the virtualization process takes care of any compromised servers without major damages. And finally, it is important to emphasize the lowering of operational costs. If the e-learning services are used only for couple of week (for providing different trainings), or couple of months (semesters in university) the savings are very high.

Looking from the other perspective, cloud computing should be more ecoefficient comparing to traditional computing. Koomey [34] suggested that there are four reasons why cloud computing is more power-efficient than the traditional in-house data center.

- Workload diversity where the resources are shared and different users request to use those resources instead of using only one institution and consuming the same power that is needed for keeping the server up and running.
- Economies of scale there are fixed costs for setting up physical data centers, in cloud computing the cost is spread more efficient since you have more users that are requesting to use the services
- **Power-management flexibility** It is easier to manage power of virtual servers than physical servcers, and in case of any failure the traffic will be redirected automatically to the other server,
- You can pick the most efficient site possible we could pick the most cost-effective location for building our data-center,

So using cloud computing you do not have to reserve hardware solutions based on the next 5 years of the growth company plan, but you can start with minimal resources and then increase the resources continually with the growth of the company. So using cloud computing, institutions can save the power consumption while still providing the servers, storage and many other network devices. Thus, it is important to emphasis that using cloud computing could be also ecoefficient [35, 36]. Whereas in [37], using cloud computing there are a lot more other technical benefits than just those aforementioned above. Starting from:

- Automation we can use different programmable APIs repeatably, without having the need to reinvent the wheel;
- Auto-scaling we can scale in and scale out our application based on current demands without any human intervention;
- More efficient proactive life-cycle Production system may be easy to clone for use as development;
- Improved test-ability Inject and automated testing at any level during the development process;
- Disaster recovery we can take the advantage of cloud that offers their servers in geo-distribution manner, you can replace your servers within a minute in case of any catastrophic failures;
- Overflow- with some click we can regulate the load balancing.

3.2 Economical Benefits of Cloud Computing

As mentioned previously the biggest benefit of cloud computing is of course the cost efficiency. Looking from the economical perspective it is cheaper to rent infrastructure than to set it up from the scratch for the businesses, when you do not have a clue how the are going to be developed in the near future. It is of course more cost effective to utilize the resources than to continually buy or sell hardware while growing or diminishing your institution. The scalability of cloud computing is the other economical key benefit. Other important economical benefits includes smaller staff, familiarity with world wide web, customization, better

marketing, security, reliability etc. Some other business benefits [37] that we can emphasize while using cloud computing are listed below:

- Zero upfront infrastructure investment Building new large scale system could cost a fortune of money and time.
- Pay-As-You-Go if your institution is growing more than you planned than you can scale your capacities just in time by self provisioning. This means that you scale your requirements only as much as you business needs and have to pay only what you use.
- More efficient resource utilization Administrators can manage the resources more effectively and efficiently by having the application request and relinquish resources on-demand.

4 e-Learning

Electronic learning is the use of technology to allow people to have access to their learning materials anytime and everywhere [38]. It is an important mode of learning today and it is becoming very popular for long life learners. Standard e-learning is not well engaging and interacting, it is usually running in particular servers and when the number of users is increasing the system performance is decreasing which is causing the system to run very slow. These problems leads students to lack of motivation to learn further. These days, e-learning is used in different education levels: training, education, virtual learning, web-based learning, continuous education, academic courses, etc.

"E" is standing for electronic learning which combines all education network activities carried out by individual or groups working online or offline through electronic devices [39]. Nowadays, people have different needs, different individual learning or teaching methodologies, who may not be satisfied with traditional learning methodologies. E-learning usually is offered as web-based open source or commercial application where the students can personalize their environment and learn on their personalized way. They can have access through a computer or mobile device without needing to sit down for hours in one place or travel in different locations for gaining the knowledge they want.

There are many content learning systems, virtual learning systems, learning management systems which mainly offer to the students the e-learning abilities.

In e-learning platforms, there are at least two factors involved: the students and the teachers. Overall, the key features of most e-learning tools or applications are how to increase collaboration, manage students, their materials, facilities, announcement, notifications, delivering web-based courses, course assessment, mock exams, displaying scores and transcripts etc. Examples of such e-learning management systems can be mentioned: Moodle [40], Blackboard [41], Sakai [42], Claroline [43] and a lot others that are being used worldwide by different education centers.

E-learning compared to traditional learning increases the learning process, changes student discussion methods, increases students motivations, changes the

role of teachers and the most important is location independent [44]. Although, analyzing from the technical perspective the main e-learning concerns are how to increase the *ility characteristics of e-learning systems, such as : manageability, durability, reusability etc., as well as continually increasing the motivation of students for continual collaboration.

By combining the characteristics of e-learning and the benefits of cloud computing infrastructure the research is concerned to focus in the improvement of those *ility issues, so the students will have motivation to learn and to surf in e-learning platforms more than they have done in standard e-learning versions.

5 The main proposal - Cloud e-Learning

Currently, different e-learning solutions are demanded since the need for education is increasing. E-learning is used widely in continuing education, training companies, academic courses etc. With the new technology evolution, e-learning should be delivered in more portable way. Usually e-learning systems are developed as distributed applications which contain systems components. Implementing the e-learning system for each education institution requires a considerable investment, since the e-learning systems are facing a challenge phase for optimizing resources, while they are continually phasing the growing of concurrent users, services, education materials and resources, rapid storage growth requirements and the last one but not the least, they are having trouble with controlling their costs.

The main idea of cloud e-learning is to focus on e-learning solutions while using the benefits of cloud computing which are offered for e-learning systems, and also emphasizing the impact of e-learning services based on cloud computing platforms. In this new technology e-learning systems will use cloud computing as an opportunity to increase the performance of the system cost-effectively. Comparing with the standard e-learning solutions, in cloud computing all elearning courses are duplicated across a number of servers and if one of servers is crashed or simply overloaded the next coming student is automatically directed to the next near server without causing any damage. These days, there are a lot of major companies that are offering services through cloud computing, especially education systems, such as: Google, Amazon, Yahoo, Microsoft etc.

The aim of cloud e-learning is to create a vision of new framework for elearning applications, which could be deployed through combined cloud types. The first part of e-learning which mainly has to do with teachers and students features can use the public cloud, whereas the administrative part which is more sensitive can use the private cloud.

While continuing with deep research and analyze, in the future we could come with result if the cloud could be used as combined and what will be the advantage of current e-learning systems? How could be involve within the elearning platform all social networks and relevant vendors who contributes to e-learning. Below, are listed a couple of research paths which could be used while continuing the research. At this stage, we are not going to decide which path we are going to follow in the future.

From an educational perspective, the main concerns of cloud e-learning will be to:

- Find main difficulties that are faced while using standard e-learning applications within an education institutions?
- How they can be adapted in our cloud e-learning framework?
- Can cloud e-learning be personalized?
- Does cloud e-learning increase the collaboration and performance?
- Does cloud e-learning do a load balancing?
- Is cloud infrastructure suitable for e-learning applications?
- What would be the appropriate PaaS framework for cloud e-learning?
- What would be the appropriate cloud deployment model for cloud e- learning?

Whereas from a technical perspective, the main research points would be focused on:

- Current cloud computing frameworks,
- Current e-learning systems,
- Identify the web based software quality factors most important in Grid and Cloud,
- Analyze parallel computing, cluster computing and other technologies,
- Analyzing Infrastructure for e-learning systems and determine the problems that are likely to be encountered when creating the proposed framework,
- Review the various components that contribute towards the cloud e-learning.

References

- 1. Richard Felder, Rebecca Brent, Understanding Students Differences, Journal of Engineering Education, 57-72, 2005
- 2. Tanya Elias, Learning Analytics: Definitions, Processes and Potential , January 2011
- 3. Christina De Simone, Richard F. Schmid, and Laura A. McEwen, Supporting the Learning Process with Collaborative Concept Mapping Using Computer-Based Communication Tools and Processes, Swets and Zeitlinger, 2001
- 4. Martin Weller, *VLE 2.0 and future directions in learning environments*, The open University, Milton Keynes, UK, LAMS conference, 2006
- Bill Buckler, A learning process model to achieve continuous improvement and innovation, Volume 3 Number 3 1996 pp. 3139, MCB University Press ISSN 0969-6474
- 6. P.Dillenbourg, Learning in the new millennium: Building new education strategies for schools, University of Geneva, EUN Conference 2000
- Florin Mihai, Andrei Stanciu, Ofelia Aleca, Changing learning environment through technology, Annales Universitatis Apulensis Series Oeconomica, 13(1), 2011
- 8. Mikael Nilsson, Matthias Palmr, Ambjrn Naeve, Semantic Web Meta-data for e-Learning Some Architectural Guidelines, Department of Numerical Analysis and Computer Science Royal Institute of Technology, Stockholm, Swede , March 2002

- 9. http://cloudofinnovation.com/category/cloud-computing/
- 10. www.thbs.com/pdfs/Comparison
- 11. http://aws.amazon.com/ec2/
- 12. Microsoft Azure, http://www.windowsazure.com
- 13. http://www.google.com/apps/intl/en/business/cloud.html
- 14. http://labs.yahoo.com,
- Sammeer Rajan, Apurva Jairath, Cloud Computing : The Fifth generation of Computing, International Conference on Communication Systems and Network Technologies, 2011
- Cloud Computing Journal, Twenty-One Experts Define Cloud Computing, Available: http://cloudcomputing.sys-con.com/node/612375/print
- 17. Derrick Kondo, Bahman Javadi, Paul Malecot, Franck Cappello, David P. Anderson, *Cost-Benefit Analysis of Cloud Computing versus Desktop Grids*, INRIA, France, UC Berkeley, USA
- Peter Mell, Timothy Grance, The NIST Definition of Cloud Computing, NIST Special Publication 800-145
- 19. Rajkumar Buyya, James Broberg, Andrzej Goscinskil, *Cloud Computing Principles* and Paradigms, by John Wiley and Sons, Inc
- Daryl C. Plummer, Thomas J. Bittman, Tom Austin, David W. Cearley, David Mitchell Smith, *Cloud Computing: Defining and Describing an Emerging Phenomenon*, Gartner 2008, [On-line]. pp. 1-9
- 21. Ian Song Tom Mainelli, Driving Productivity with the Anywhere, Anytime Virtual Desktop, IDC Analyze the future
- 22. Rackspace, http://www.rackspace.com
- 23. IBM, http://www.ibm.com/cloud-computing/us/en/
- 24. Google App Engine, https://developers.google.com/appengine/
- 25. Google Docs, https://drive.google.com/
- 26. Vijay Sarathy, Purnendu Narayan, Roa Mikkilinwni, Next generation Cloud Computing Architecture, Los Altos, CA, www.kawaobjects.com/resources/PID1258479.pdf
- 27. Armbrust M., Fox A., Griffith R., Joseph D., Katz R., Konwinski A., Lee H., Patterson D., Rabkin A., Stoica I., Zaharia M., *Above the Clouds: A Berkeley View of Cloud Computing*, Technical report No: UCB/EECS-2009, UC Berkeley Reliable Adaptive
- 28. Roger Clarke, User Requirements for Cloud Computing Architecture, 10th IEEE/ACM International conference 2010, in Cluster, Cloud and Grid Computing
- O. Ingthorson, Concepts and Technologies for cloud computing and Software as a Service, 2009
- Claudio T., Ricardo A.Joaquim S.P, Tiago B., User Provided Cloud Computing, 10th IEEE/ACM International conference - 2010, in Cluster, Cloud and Grid Computing
- GTSI Solutions, Cloud Computing, Building a Framework for Successful Transition, http://www.gtsi.com/cms/documents/White-Papers/Cloud-Computing.pdf, Accessed: [August 18, 2011]
- 32. http://cioresearchcenter.com/2010/12/107/
- 33. T O'Reilly, Is Google App Engine a Lock-in Play , 2008, http://radar.oreilly.com/2008/04/is-google-app-engine-a-lockin.html
- Jonathan Koomey, Advantages of Cloud Computing, February 2012, http://www.knowcloudcomputing.blogspot.com/2012/02/advantages-of-cloudcomputingaccording.html/

- William Fellows, Andy Lawrence, Eco efficient IT and the greening of the cloud, presented at the 2nd Int. Symposium - Data Center Efficient and Green Enterprises , New York, USA, May 2010
- 36. Expert Group Report, The future of cloud computing, opportunities for European cloud computing beyond 2010, European Communities, 2010, http://cordis.europa .eu/fp7/ict/ssai/docs/cloud-report- final.pdf
- 37. Jinesh Varia, Amazon Web Services Architecting the Cloud: Best Practices, 2010
- Goldman Sachs, What is Electronic Learning?, [On-line]. pp. 3-19, http://www.mup.com.au/uploads/files/pdf/978-0-522-85130-4.pdf
- Som Naidu, E-learning, A Guidebook of Principles, Procedures and Practices, 2nd ed., CEMCA, 2006, pp.1-100
- 40. *Moodle*, http://moodle.org/
- 41. Blackboard, http://www.blackboard.com/
- 42. http://www.sakaiproject.org/
- 43. http://www.claroline.net/?lang=en
- 44. David Rashty, Traditional Learning vs eLearning Turrof, 1995
- 45. Thomas B. Winans, John S. Brown Cloud Computing A Collection of Working Papers, May, 2009
- Heather Clancy, 4 power advantages of cloud computing, http://www.zdnet.com/blog/green/4-power-advantages-of-cloudcomputing/11770
- 47. Ted Alford, Gwen Morton, The Economics of Cloud Computing, Addressing the benefits of Infrastructure in the cloud, Booz Allen Hamilton Inc, 2010
- 48. Clark, R. C., and Mayer, R. E. *e-learning and the science of instruction*, San Francisco: Jossey-Bass.
- Bo Dong, Qinghua Zheng, Mu Qiao, Jian Shu, and Jie Yang, BlueSky Cloud Framework: An E-Learning Framework Embracing Cloud Computing, Springer-Verlag Berlin Heidelberg 2009
- 50. ZDNet, Cloudy Concept: IaaS, PaaS, SaaS, http://www.zdnet.co.uk

Addressing the Challenge of Application Portability in Cloud Platforms

Fotis Gonidis, Iraklis Paraskakis, Dimitrios Kourtesis

South-East European Research Centre (SEERC), City College – International Faculty of the University of Sheffield, 24 Proxenou Koromila Street, 54622 Thessaloniki, Greece {fgonidis, iparaskakis, dkourtesis}@seerc.org

Abstract. Cloud computing is a relatively new paradigm that promises to revolutionize the way IT services are provided. There are multiple benefits that companies can gain from cloud computing. However, there still remain a number of issues to be solved before this new computing paradigm is widely adopted. In this paper we introduce the issues of portability and interoperability and we focus on the cross-platform portability of applications. We present some high level approaches and existing work that try to address this issue. Finally we briefly propose some future research directions towards investigating how to improve the portability of applications.

Keywords: Cloud Computing, Portability, Interoperability

1 Introduction

Cloud computing is a relatively new paradigm, where computing is offered as a utility, and has the potential to transform a large part of the IT industry [1]. According to the US National Institute of Standards and Technology (NIST), cloud computing is a "model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [2].

Due to the fact that cloud computing refers to provisioning of heterogeneous resources, ranging from hardware (storage, processing power) to software (software development platforms, applications), it has been further decomposed into various service layers. According to NIST there are three service layers [3] which are shown in Figure 1. The bottom service layer is the Infrastructure as a Service (IaaS). In IaaS, basic computing resources like storage, processing and networking are provisioned. Amazon is a major IaaS vendor, providing computing resources via virtual machines. The middle service layer is the Platform as a Service (PaaS). In PaaS the consumers are given access to software tools, APIs and programming languages in order to develop and deploy their applications. Example of a PaaS provider is Google, which via Google App Engine provides developers with programming runtime environments, storage options, custom services, deployment capabilities etc. The top

level service layer is the Software as a Service (SaaS). In this service layer, complete and ready-to-be-used software applications are provisioned. SalesForce is a major SaaS provider offering Customer Relationship Management (CRM) Software.

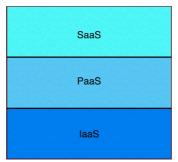


Fig 1. The three service layers of cloud computing

Cloud consumers, whether they are acting on IaaS, PaaS, or SaaS layer can benefit a lot from this computing paradigm. Using cloud computing, consumers have ubiquitous access to their data and applications from all over the world. The only requirement is access to the internet. Consumers no longer need to worry about installing, maintaining and upgrading the software and/or hardware that they use. This burden has now moved entirely to the cloud providers. Another benefit comes from the characteristic of cloud computing to rapidly provision and release resources. Consumers can have the resources that they need available at any given time and pay only for what they use. Therefore they don't need to buy upfront software or hardware that they may never use.

While cloud computing is rapidly emerging and offers several benefits for consumers, there are still a number of issues to be addressed before it becomes widely adopted by companies and organizations. Among others, important issues are portability and interoperability across cloud platforms. As Figure 2 shows, the first refers to the ability to move a service from one cloud provider to another, while the latter one refers to the ability of two services from different clouds to exchange information.



Fig 2. Left: a component is moved from one environment to another (portability). Right: two components of different cloud environments are collaborating (interoperability).

As will be presented in the next section, portability and interoperability affect all three service layers (IaaS/PaaS/SaaS). In this paper we are particularly interested in

the aspect of portability within the context of PaaS, namely the feasibility of a cloud application to be ported across different platforms. Major cloud platform providers like, Google App Engine and Microsoft Azure are currently using proprietary technologies and/or techniques that prevent the portability of cloud applications. This issue plays an important role in the wide adoption of cloud computing as it may discourage consumers from using cloud platform services.

The aim of this paper is to introduce portability and interoperability as concepts, discuss their relationship, and present some high level approaches for addressing the challenge of portability in the context of PaaS. The presented approaches vary from defining common standards, creating APIs, abstract/wrap proprietary ones, to developing open source platforms. Our research goal is to explore how to improve the portability of cloud applications, and the first objective towards this goal is to understand how platform-independence can be enhanced by employing Model Driven Engineering (MDE) and Ontologies.

The rest of the document is organized as follows. In the next section the terms portability and interoperability are introduced and it is explained why they are important issues for the wide adoption of cloud computing. Next, we focus on the issue of portability in the context of the PaaS layer. We discuss high level approaches towards enabling portable cloud applications, followed by an overview of existing work in this field. Finally, in section 4 we present some future research directions.

2 Portability and Interoperability Issues in Cloud Computing

Prior to focusing on the aspect of portability in the context of PaaS, it is necessary to introduce the concept of portability in the context of cloud computing and explain how it differs from that of interoperability.

2.1 Portability in Cloud Computing

NIST refers to portability as the ability "of prospective cloud computing customers to move their data or applications across multiple cloud environments at low cost and minimal disruption" and particularly to system portability as "the ability to migrate a fully-stopped Virtual Machine (VM) instance or a machine image from one provider to another provider" [3]. Two important characteristics of portability can be extracted from this definition. Firstly, the move from one cloud to another should be achieved at the lowest possible cost, effort, and time. Secondly, portability refers to the ability to move any component of any of the three service layers across cloud platforms.

According to Cloud Security Alliance (CSA) [4], there are different portability requirements at the three different cloud service layers. In IaaS, the requirement is to be able to easily port the Virtual Machines (VMs) and data from one vendor to another. For instance, a company or organization operating several VMs in one cloud infrastructure provider should be able to easily move them to another provider. In PaaS, the requirement is to be able to deploy applications across different platforms. For example, if a developer creates and deploys an application on a certain cloud platform, it should be feasible for the application to be ported to a different platform,

with a minimal set of changes, if any. Finally, in SaaS, the requirement, when switching from one software application to another is to be able to extract the data from the first and upload it to the second. For example, if a company uses a CRM application provided by a cloud vendor and decides to switch to a different offering, it is important that all customer data can directly be loaded and processed by the new CRM application.

Consumers, whether they interact with the IaaS, PaaS or SaaS layer need to be able to easily change between cloud providers and be free to choose the one that better serves their needs in terms of quality and/or cost. The ability of consumers to easily migrate from one cloud service provider to another is even more critical in case a cloud provider's operation is unexpectedly terminated. A real example to illustrate this argument is the case of Coghead [5] – an online application development platform supporting the development and hosting of data-driven applications. The company had managed to attract several hundreds of developers before it suddenly announced that it would stop operating, calling all customers to export the data stored in their applications, but not giving them the option to port the applications themselves.

Ensuring portability across cloud providers would eliminate the vendor lock-in problem [6] and would allow consumers to switch between vendors according to their needs. In turn, this would increase consumers' trust towards cloud computing and public cloud services.

2.2 Interoperability in Cloud Computing

In the IEEE glossary [7], interoperability is defined as "the ability of two or more systems or components to exchange information and to use the information that has been exchanged". According to Petcu [8], one can find several definitions of cloud interoperability in the literature. For example, interoperability has been defined as the ability to "abstract the programmatic differences from one cloud to another", the ability to "translate between the abstractions supported by different clouds", to "flexibly run applications locally or in the cloud or in a combination", or to "use same management tools, server images, software in multiple clouds".

Interoperability affects all three service layers and there are specific requirements in each one of them. In IaaS, interoperability may refer to the ability of a client to seamlessly use infrastructure resources from different vendors through a common management API [9]. For instance, a consumer could be able to perform the same set of operations on VMs from different providers (e.g. start, stop or delete) without creating a different client for each of the providers. In PaaS, developers may need to use tools, libraries or APIs coming from different PaaS providers to create their applications. For example a cloud application may be composed of various cloud services coming from different cloud vendors. Finally, in SaaS, interoperability lies in the ability of different applications to exchange information. For example as Dillon et al. [10] mention, a company might want to outsource the email service to Google and the Human Resource Management (HRM) service to SalesForce. This means that the data format in the e-mail system (e.g. calendar, address book) needs to be compatible with the HRM service. Lack of interoperability puts a barrier to using and combining solutions from different providers, but also to allowing on-site systems to collaborate and exchange information with cloud services. Enhancing interoperability among cloud providers would enable two ways of collaborations. Firstly, services from various cloud providers could be seamlessly combined to provide best of breed solutions with respect to quality, price, or features. Secondly, companies would be able to "push" to the cloud only part of their services while keeping the most critical ones on–site.

2.3 Relationship Between Portability and Interoperability

In 2.1 and 2.2 portability and interoperability were presented as distinct notions. Nevertheless, it is not uncommon for the term "interoperability" to be used for denoting both concepts. As noted by Petcu [8], authors sometimes define interoperability as the ability to "move applications from one environment to another or run in multiple clouds", or the ability to "move services, processes, workloads, and data between clouds". According to the definition we provided in section 2.1, this is clearly a case of portability. In Dowell et al. [18], portability is referred to as a special kind of interoperability challenge. Seen from this perspective, the inconsistent use of the term "interoperability" to denote both concepts can be justified.

As proposed by Petcu [8], it is perhaps useful to think in terms of vertical and horizontal interoperability. Horizontal interoperability can be defined as the ability of two cloud services of the same service layer (IaaS/PaaS/SaaS) to communicate with each other – a definition consistent with our notion of interoperability as discussed in section 2.2. On the other hand, vertical interoperability can be defined as the ability of a cloud service to be deployed on a cloud service of a lower service layer. For example, allowing a SaaS application to be deployed on various PaaS offerings. This definition is consistent with our notion of portability as discussed in section 2.1, i.e. the ability of an application to be ported across various cloud platforms.

3 Portability in the Context of Platform as a Service

As mentioned in section 2.1, there are portability requirements across all three service layers (IaaS/PaaS/SaaS). The scope of this research, however, is to explore the issue of application portability at the level of PaaS.

3.1 Cloud Portability Issues at Platform Level

Cloud platforms promise to ease and speed up the application development cycle by offering a complete set of tools for developing, deploying, hosting and maintaining the application. There is presently a broad choice of providers of PaaS, such as Google App Engine, Microsoft Azure, and Force.com, who offer a wide range of services for application development, including file and data storage, messaging, queuing, workload management, analytics etc [20]. The combination of the large number of platform offerings and the variety of cloud services contributes to the

portability challenge, as it makes moving an application across cloud platforms even more complex.

The above statement becomes clearer if we consider that the cloud services, according to each provider, may use different technologies and be offered to clients by various proprietary APIs. For example, Microsoft provides the SQL Azure database for data storage, while Google App Engine provides, among others, the App Engine DataStore. On top of that, application portability may be hindered by the fact that certain cloud services offered by one platform are not available in another. For example, the mailing service offered by Google App Engine may not be offered by any other provider. Thus, the heterogeneity of today's platform offerings contributes significantly to the challenge of application portability.

3.2 General Approaches for Addressing Cloud Portability in PaaS

There are some generic approaches and strategies that could be adopted in order to tackle the issues outlined in section 3.1, and eventually ease application portability across platforms.

One obvious approach is the definition of common set of standards for PaaS offerings. The adoption of such standards by all cloud providers would enable developers to create their applications independently of specific platform environments and then deploy them to the cloud platform of their choice. This set of standards could include a standardized API to access the service offered by the platform, standard formats for representing file structures, standard data stores, etc.

Standardization seems to be a very efficient approach to achieve cloud portability. However, for reasons not necessarily related to technology, it is very difficult for all cloud platforms to eventually agree on a common set of standards. All major cloud vendors use proprietary APIs and file formats as a way of locking-in customers to their services. The effort required to re-engineer an application in order for it to be ported to another platform is discouraging customers to move. In addition, a set of common standards would prevent platform providers from offering the special, platform-specific features that allow vendors to differentiate from their competitors.

Another approach towards achieving portability between platforms is intermediation. That is, introducing an intermediate layer that decouples application development from specific platform APIs and supported formats. In this case developers create their applications using an intermediate API which is platform agnostic and can "hide" or "wrap" the proprietary APIs of particular vendors. The intermediate layer prevents developers from being bound to specific programming languages, file formats or data stores. For example an application could be developed in a language-independent manner, and later on, through model transformations, be translated into the particular programming language supported by a PaaS provider (such as Java, Python or C#), or the database query language particular to a platform (e.g. Microsoft SQL or MySQL).

In this case, of introducing an intermediate layer for decoupling application development from specific platforms, no consensus by platforms vendors is required. However, the challenging part here is to develop the translation rules and the model transformations between the intermediate layer and each platform vendor specifically.

3.3 Existing Works Addressing Cloud Portability at Platform Level

There are several existing works that try to address the issue of portability across cloud platforms, by adopting one or combining the two approaches presented in section 3.2.

mOSAIC. mOSAIC is a framework that promises to ease application portability 1) across platforms by providing a set of APIs that are independent of vendors [12], [13]. At design-time, developers are using these APIs to create applications that consist of multiple cloud components, each one performing a certain function. A cloud component can, for example, be a Java application. At this point the application is not bound to any specific platform. Then, at runtime, the mOSAIC platform decomposes the application into the various cloud components and deploys each one on the cloud platform that provides the best implementation for the cloud component's functionality. Communication between components deployed in different platforms is achieved via cloud based message queues technologies [8]. The code for connecting the application components to a concrete platform provider is generated by mOSAIC. Therefore developers can focus on developing their applications in a platform-neutral manner, and later on, they can decide on which cloud provider they wish to deploy them. The mOSAIC API acts as an intermediary layer between the developers and the actual cloud platforms, and developers do not have to use proprietary APIs of the target could platforms. Therefore mOSAIC could be classified as an intermediation approach that tries to decouple application development from particular platform technologies.

An example application, as described in Petcu et al [13], could be a check out service for buying products online. The application could be split into four core operations: a) retrieving user payment details and product list, b) calculating the total amount to be charged to the user's credit card, c) charging the credit card by contacting the bank, d) saving the transaction details. At development time, each of these operations is assigned to a cloud component, and during run-time each component can be deployed on a platform that best performs the operation.

- 2) Open Cloud Computing Interface (OCCI): OCCI is a set of specifications that allow the development of tools for performing common cloud tasks like deployment, autonomic scaling and monitoring across different cloud service providers. It offers an API that is supported by various cloud computing stacks, like Open Eucalyptus, OpenNebula, and OpenStack. Therefore, OCCI could be classified as a standardization approach. It should be mentioned that OCCI started as an API for managing cloud infrastructure. However, as it is stated in its official website, the Open Cloud Computing Interface will eventually also serve other layers in addition to IaaS, i.e. PaaS and SaaS [19].
- 3) PaaS Semantic Interoperability Framework (PSIF). The PSIF framework proposed by Loutas et al. [14] aims at modelling semantic interoperability conflicts that may occur during migration or deployment of an application on a cloud platform. The framework is structured according to 3 dimensions, (i) the different architectural entities in a PaaS environment, (ii) the type of semantics of a PaaS entity's description i.e. functional, non-functional and execution

semantics, (iii) and the level at which semantic conflicts occur, i.e. the level of the information model and the level of data. Semantic conflicts are identified and classified according to these 3 dimensions.

Loutas et al. [15], provide two examples of the framework's operation. In the first example, an application is ported from one platform to another. A conflict arises when the application is trying to connect to a database, because the two platforms use different function calls (e.g. "connect a db" vs. "insert a db"). This semantic conflict occurs due to differences in the definitions of the management interfaces of the two platforms and specifically due to the way their functional semantics are modelled. The conflict is raised at the data level, since it is caused by different naming of the same functionality. Another semantic conflict may occur due to differences in the modelling of the PaaS offerings. For example, one provider uses a field "programming language" to describe both the language and the version, e.g. Java 1.6, while another platform offering uses two different fields. In this case the same term has different meaning in each platform. The conflict is raised due to differences in the semantic models of the PaaS offerings and specifically to the way the non-functional semantics are modelled. The conflict occurs at the information model level, since it is caused by different logical representation of the same information. In a similar way, other semantic conflicts which may occur while moving applications across platforms are classified

Having modelled in detail the fundamental PaaS entities in a particular PaaS offering, a semantic layer will be implemented to provide a PaaS Offering Model and an Application Model for the common description of available PaaS offerings [15]. PaaS providers will be able to publish their offerings based on these common models. By letting providers adopt a common model for their offerings the application portability across the platforms will be enhanced. According to our understanding and the available literature on the PSIF, we could classify this work as an approach of defining a set of common standards.

4) SimpleCloud. SimpleCloud is an API that allows developers to use storage services independently of particular cloud platforms. Among others, it offers two key services: (i) File Storage Service and (ii) Document Storage Service. The File Storage Service allows for performing operations on files such as storing, reading, deleting, copying, storing metadata, etc. It does so by providing so-called "storage service adapters" that allow developers to access storage services from Amazon, Microsoft Azure, Rackspace and others, using the same application code. The Document Storage Service abstracts the interfaces of all major document databases, again allowing developers to access different providers through a single API. SimpleCloud is supported by IBM, Microsoft, Rackspace, GoGrid, and several other cloud service providers. By offering an API for storing data which abstracts/hides all proprietary ones, SimpleCloud can be considered as an intermediate layer for decoupling applications from directly accessing the storage mechanisms of specific platforms.

4 Future Research Directions

This research work is still at an initial stage and is carried out in the context of the first author's doctoral research. As it was already mentioned in section 3, the research scope is to explore and propose a framework for improving the cross-platform development and deployment of cloud applications.

Today, there is a wide range of available platform offerings that a cloud developer can choose from. Each platform may have unique characteristics and use certain technologies, tools, APIs etc. Moreover the provided services may vary significantly across platforms. Therefore, it is not feasible to achieve a single generic solution that could enable applications to be ported across all available cloud platforms without any modifications. A realistic approach would be to focus on a certain set of platform offerings, services, technologies, as in the previously mentioned cases of SimpleCloud, which supports PHP and focuses on abstracting cloud storage mechanisms, or mOSAIC, whose scope is to support Java and Python applications.

Therefore the major research question at this stage of our work is to define and map the area within the context of PaaS that this research will focus on. In order to answer this question there are several sub-questions that need to be addressed first. Firstly, a survey of the available platform offerings needs to be performed. Afterwards, these platforms need to be examined and analyzed in order to extract the platform components and characteristics of each one of them. Then it would be feasible to examine what are the specific conflicts, and at which level they occur while porting an application across platforms (e.g. an obvious one is the programming language while moving from Microsoft Azure to Google App Engine, since the first primarily supports C# while the latter supports Java and Python). After forming a clear idea about the various platform offerings and the characteristics of the cloud platform, we will able to narrow down our research effort to a specific set of platform vendors and platform characteristics that are prone to conflicts when porting an application. At the same time, an extensive survey of related work in the field should be performed. An analysis of this work should be followed to identify their exact contribution and also their limitations. This will improve our understanding of the field and enable us to better define our research direction and focus.

In section 3.2 we presented two generic approaches for enhancing application portability. The first one involved the definition of a common set of standards that need to be adopted by platform vendors in order to provide uniform services. This approach however, presupposes the agreement of major platform providers and organizations, and is therefore beyond the scope of this research work. Thus, our intention is to explore the second approach which proposes an intermediary layer for decoupling the development of applications from specific proprietary technologies. This approach introduces the notion of platform-independence, where the application is initially developed in a platform-agnostic way. A research area and domain of technology which is deemed highly probable to have a positive contribution towards achieving platform independence is Model Driven Engineering (MDE).

MDE allows developers to create applications independently of a target platform, by first creating a platform-agnostic model. This intermediate model "hides" specific characteristics of platforms that could cause conflicts when deploying an application. The intermediate model allows developers to avoid binding applications to particular programming languages or to a specific database query languages. Subsequently, by performing automated model transformations, the intermediate models can be translated into platform-specific ones. The key characteristic of MDE, which is to allow developers to work with abstract models, has several benefits. According to Esparza-Peidro and Munoz-Escoi [21], it allows developers to focus on their applications and ignore minor details linked to specific platforms. They can avoid low-level and error-prone platform features and can more efficiently communicate with each other. Therefore the overall design and implementation process can be dramatically improved.

A most relevant benefit in relation to our particular research focus is the abstract nature of models, which has the potential to improve cross-platform development and deployment of applications by decoupling the development from specific platform technology. For example, we could think of an application that requires the use of a noSQL database. OpenShift [23], the PaaS offering from Red Hat, supports, among others, mongoDB. Google App Engine [24] on the other hand provides a noSQL database called App Engine Datastore. A conflict is raised when the application developer needs to move the application from one platform to another. MDE, in this case, could allow the developer to model the database as an abstract model. Later on, through automatic model transformations, the abstract model could be mapped on a specific database implementation according to the target platform. Therefore we intend to explore what the contribution of using MDE techniques could be in creating cloud applications that will be more independent of the targeted platforms.

On top of using MDE, in this research we are particularly interested in exploring whether involving ontologies could contribute to improving cloud portability at the level of PaaS. Ontologies, according to Gruber [22], are a "formal explicit specification of a shared conceptualization". In other words, ontologies can provide a common vocabulary and understanding over a particularly domain. Specifically in PaaS layer, the use of a common acceptable ontology can lead to more uniform offerings from PaaS providers' side. Therefore application portability can be improved. The PSIF, presented in 3.3, is inspired from this notion and aspires to provide such an ontological model for PaaS offerings. This is one approach of how ontologies could contribute to the issue of cloud application portability. Another interesting research direction to explore is how ontologies can be combined with MDE and what the potential benefits of such a combination could be in the field of application portability.

Gasevic et al. [16], as well as Happel and Seedorf [17], support the view that combining ontologies with model-driven engineering during software application development has many benefits. Ontologies provide support for logical inference and if combined with software models, such as UML, they can enable reasoning over the models used in model-driven engineering. The ability to perform reasoning and inferring new knowledge from software models motivate us to explore the benefits of combining MDE and ontologies.

In the area of cloud platforms, Ranabahu and Seth exploit the combination of MDE and ontologies. MDE is used in order to abstract platform specific details while ontologies enrich the models with more information. For example, in a demo application they define a component to perform a certain action that needs to be secured with the "ssl" protocol. In order to add this piece of information, they semantically annotate the model using an available security profile ontology.

Therefore it is obvious that it is not the first time that ontologies are used to tackle the issue of portability in cloud applications. However to the best of our knowledge, the combination of MDE and ontologies for enabling cloud portability at PaaS model is an area that has not yet been thoroughly explored. This fact further motivates our research direction.

5 Conclusions

In this paper, we introduced the challenges of portability and interoperability in cloud computing and provided a brief overview of portability and interoperability issues specific to each of the three cloud service layers (IaaS, PaaS, and SaaS). The focus of our ongoing research work is to address the challenge of application portability in the context of PaaS, i.e. exploring ways to enable applications to be deployed across various cloud application platforms. In this context, we gave a high level description of the factors that may have an impact on application portability, such as proprietary APIs and differences in the functionality across different platform providers. We have also described the two high level strategies that can be employed to address the challenge of portability: standardization and intermediation. Standardization addresses cross-platform portability through the adoption of common standards by cloud providers. Alternatively, intermediation enables developers to create applications independently of a specific platform and then bind them to particular target platforms through some form of automatic translation. In the next steps of our research we will focus on exploring how to improve application portability by employing an intermediation strategy that combines elements of model-driven engineering and ontological modelling.

References

- M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.H. Katz, A. Konwinski, G. Lee, D.A. Patterson, A. Rabkin, I. Stoica and Matei Zaharia, "Above the Clouds: A Berkeley View of Cloud Computing", UC Berkeley Reliable Adaptive Distributed Systems Laboratory, 2009
- 2. P. Mell and T. Grance, "The NIST definition of cloud computing", National Institute of Standards and Technology, Gaithersburg, 2011.
- Fang Liu, Jin Tong, Jian Mao, Robert Bohn, John Messina, Lee Badger and Dawn Leaf, "NIST Cloud Computing Reference Architecture", National Institute of Standards and Technology, Gaithersburg, 2011.
- Cloud Security Alliance, "Security Guidance for Critical Areas of Focus in Cloud Computing V2.1", Cloud Security Alliance, 2009.
- J. Kincaid, "Coghead Grinds To A Halt, Heads To The Deadpool", (techcrunch), [online] 2009, <u>http://techcrunch.com/2009/02/18/coghead-grinds-to-a-halt-heads-to-the-deadpool/</u> (Accessed: 2012)
- Neal Leavitt, "Is Cloud Computing Really Ready for Prime Time?", *Computer*, vol. 42, no. 1, pp. 15–20, 2009.

- "IEEE Standard Glossary", [online], http://www.ieee.org/education_careers/education/standards/standards_glossary.html, (Accessed: 2012).
- 8. D. Petcu, "Portability and interoperability between clouds: challenges and case study", in *Proceedings of the 4th European conference on towards a service-based internet*, Poznan, 2011, pp. 62–74.
- 9. DMTF, "Interoperable Clouds A White Paper from the Open Cloud Standards Incubator", Distributed Management Task Force Inc., Portland, DSP-IS0101, Nov. 11, 2009.
- T. Dillon, Chen Wu, and E. Chang, "Cloud Computing: Issues and Challenges", in 2010 24th IEEE International Conference on Advanced Information Networking and Applications (AINA), Perth, 2010, pp. 27–33.
- B. P. Rimal, Eunmi Choi, and I. Lumb, "A Taxonomy and Survey of Cloud Computing Systems", in *Fifth International Joint Conference on INC, IMS and IDC, 2009. NCM '09*, Seoul, 2009, pp. 44–51.
- F. Moscato, R. Aversa, B. D. Martino, T.-F. Fortis, and V. Munteanu, "An Analysis of mOSAIC ontology for Cloud Resources annotation", in *FedCSIS*, 2011, Szczecin Poland pp. 973–980.
- 13. D. Petcu, G. Macariu, S. Panica, and C. Crăciun, "Portable Cloud applications—From theory to practice", *Future Generation Computer Systems*, 2012.
- N. Loutas, E. Kamateri, and K. Tarabanis, "A Semantic Interoperability Framework for Cloud Platform as a Service", in 2011 IEEE Third International Conference on Cloud Computing Technology and Science (CloudCom), Athens, 2011, pp. 280–287.
- N. Loutas, E. Kamateri, and K. Tarabanis, "Cloud Semantic Interoperability Framework", Cloud4SOA, D1.2, 2011.
- 16. D. Gasevic and N. Kaviani, "Ontologies and software engineering." in *Handbook on Ontologies*. Springer Berlin Heidelberg, 2009, pp. 593-617
- 17. H.J Happel, S. Seedorf, "Applications of ontologies in software engineering", in *Proc. 2nd InternationalWorkshop on SemanticWeb Enabled Software Engineering (SWESE 2006)*, Athens, USA, 2006
- S Dowell, A. Barreto III, J.B Michel, and M.T. Shing, "Cloud to Cloud Interoperability", in Proceedings of the 2011 6th International Conference on Systems Engineering, Jun 27-30 2011, Albuquerque, New Mexico, pp. 258-263
- 19. Open Cloud Computing Interface, OCCI, [online], <u>http://occi-wg.org/</u> (Accessed: 2012)
- 20. S. Charrington, "Don't Pass on PaaS in 2010," (ebizo), [online] 2010, <u>http://www.ebizq.net/topics/cloud_computing/features/12279.html?page=2,</u> (Accessed: 2012)
- J. Esparza-Peidro, F.D. Munoz-Escoi, "Towards the Next Generation of Model-Driven Cloud Platforms", Instituto Universitario Mixto Tecnologico de Informatica, Valencia, TR-ITI-SIDI-2011/001, 2011
- T. Gruber, "A translation approach to portable ontologies." Knowledge Acquisition, vol.5, no.2, pp.199-220, 1993
- 23. OpenShift, [online], <u>https://openshift.redhat.com/app/</u> (Accessed: 2012)
- 24. Google App Engine, [online], <u>https://developers.google.com/appengine/</u> (Accessed: 2012)

Security Assessment of OpenStack Open Source Cloud Solution

Sasko Ristov, Marjan Gusev, and Magdalena Kostoska

Ss. Cyril and Methodius University / Faculty of Information Sciences and Computer Engineering, Rugjer Boshkovik 16, PO Box 393, 1000 Skopje, Macedonia {Sashko.Ristov,Marjan.Gushev,Magdalena.Kostoska}@finki.ukim.mk

Abstract. Cloud paradigm opens a series of new security challenges for customers. The decision to migrate the services from on-premise resources to public clouds relies on the security measures that cloud service provider takes, but more important is the trustworthiness between the cloud service provider and the cloud customer. Another solution is to create a private cloud in which the customer is fully responsible for security challenges.

In this paper we assess the security aspects of the OpenStack cloud software solution. Although OpenStack offers a very scalable and flexible platform for IaaS and a lot of security measures are taken, still we found security in-compliance with several ISO 27001:2005 controls and control objectives that directly depend on the cloud software solution. We also evaluate and compare the compliance of other common open source cloud solutions as offer the customers a possibility to build their own private clouds or even a hybrid cloud.

Keywords: Cloud Architecture, Cloud Computing Security, ISO 27001:2005, Open source, OpenStack

1 Introduction

There are a lot of cloud service providers (CSPs) offering various cloud solutions, such as Amazon's AWS [1], Salesforce's Sales Cloud [21], Google's App Engine [8] and Cloud Storage [9], Microsoft Azure [13] and Live [12], VMware's vCloud [23] and many others. Voras et al. [24] devise a set of criteria to evaluate and compare most common open source IaaS cloud solutions.

Cloud computing is the newest, the fifth generation of computing after Mainframe, Personal Computer, Client-Server Computing, and the Web [17]. Each new service provides better opportunities and performance. The quality managers put their efforts in service's quality, easy usage, less price, good advertising, availability, design, etc. Business managers know that risks exist in spite of all the benefits of every new technology or business model offers. Also, many issues like regulatory violation, security, trust and privacy appear [19].

Cloud computing arises new security challenges besides the existing ones. Ristov et al. [18] give a comprehensive analysis for business information system security in cloud computing for each cloud service layer IaaS, PaaS and SaaS compared to traditional on-premise solution. [20] presents a new methodology for security evaluation in cloud computing. The main goal of this paper is to assess the security aspects of the OpenStack solution by analyzing and comparing it to the other open source cloud solutions. We focus on OpenStack because more than 75 leading companies participate in its development including Cisco, Citrix, Dell, Intel and Microsoft.

The paper is organized as follows. Section 2 briefly describes most common open source cloud softwares. In Section 3 we assess the information security issues in open source cloud softwares, focusing on OpenStack. We emphasis the results of the security assessment of OpenStack cloud software solution as security pros and cons in Section 5. Section 4 evaluates and compares the open source cloud security challenges according to ISO 27001:2005 controls and control objectives that depend of the cloud solution. Finally we conclude our work in Section 6.

2 The Open Source Cloud Solutions

Despite the commercial clouds and their services, there are many open source cloud solutions that provide the customers to develop their own private cloud, especially IaaS cloud service layer, such as well known OpenStack [16], Eucalyptus [7], OpenNebula [15], and CloudStack [6]. There are open cloud-computing research testbed designed to support research into the design, provisioning, and management of services at a global, multi-datacenter scale, such as Opencyrrus [5]. It is presented in [2] [4].

Almost all open source cloud solutions have the same architecture. Eucalyptus has a similar architecture as OpenStack [14]. Each solution consists of several different parts and all have two main parts: The cloud controller and Nodes. The former controls the system, network, schedules the instances and is the administrator interface. The latter runs the instances of the virtual machines and uses the available hardware resources. Another important part of each solution is the network management, i.e. delimiting and managing the public and private network.

Almost all open source cloud solutions provide interfaces to commercial cloud services Amazon's EC2 and S3, and Google's App Engine. They also allow users to build not only their own private cloud, but also a hybrid cloud. The open source solutions give the freedom to customer to choose the hardware and software vendor. All main hypervisors and operating systems are supported.

In this paper we focus on OpenStack Cloud Software. It is a open source cloud software that builds both public and private clouds. *OpenStack: The Essex Release* is the latest (the fifth) released version.

3 OpenStack Security Assessment

In this section we assess the information security challenges in open source cloud softwares focusing on OpenStack cloud solution. Tsai et al. address the system security, networking security, user authentication, and data security as information security issue of cloud computing in [22]. In this paper we add and analyze additional security issues in the security assessment.

3.1 User Access Management

User access management is based on *Role Based Access Control* (RBAC) in OpenStack. Each role has predefined set of permitted operations. Then roles are assigned to each user. The roles can be assigned when the user is created or editing the existing user profile.

Administrator is a project based role and has a privilege to add or remove an instance, to remove an image and to add a key. *IT security* is global role that permits role holders to quarantine instances. *Project Manager* is default role for project owners and permits the role holders to add or revoke roles to users of the project, to manage network and all the privileges that Administrator role has. *Network Administrator* role allows particular user to allocate and assign publicly accessible IP addresses, and to create and modify firewall rules. *Developer* role is a global role that is assigned to users by default and user assigned with this role can create and download keys.

OpenNebula has the similar RBAC policy. Eucalyptus and CloudStack have more features such as LDAP integration and X.509 certificates.

3.2 Network Access Management

OpenStack Nova supports three kinds of networks: Flat Network Manager, Flat DHCP Network Manager, and VLAN Network Manager. The first two kinds of network assign the IP addresses of the subnet that is specified by the user with network administrator role. Therefore this control objective is not satisfied with these two kinds of networks.

In VLAN Network mode, Nova creates a VLAN and bridge for each project. Each project gets a range of private IP addresses that are only accessible from inside the VLAN. Each project gets its own VLAN, Linux networking bridge, and subnet in this mode.

A special VPN instance needs to be created in order for a user to access the instances of the virtual machine in their project. Certificate and key are generated to access the VPN which provides private network segment for instances of the project.

Berger et al. [3] develop a trusted virtual datacenter (TVDc) technology to address the need for strong isolation and integrity guarantees in virtualized, cloud computing environments. It clearly splits the management responsibilities between the cloud and tenant administrators.

All analyzed open source solutions have integrated firewall rules within the controller.

3.3 Operating System Access Control

Operating systems of the instances are controlled by the users. Those on the OpenStack physical servers are managed by the system administrator. The volumes, image and storage data are secured with project keys.

OpenNebula offers three authentication for the images: user-password scheme, x509 certificates based authentication and EC2. CloudStack has LDAP based authentication and Eucalyptus has LDAP and secret keys based authentication.

3.4 Application and Information Access Control

OpenStack has the application Dashboard for image, volume and instance management. The access is controlled via username and password. Khan et al. [11] propose and implement OpenID-Authentication-as-a-Service APIs that allows users to use their OpenID Identifiers to log into the OpenStack Dashboard.

Other analyzed open source solutions have integrated LDAP and X.509 certificates.

3.5 Mobile Computing and Teleworking

OpenStack ensures information security when using mobile computing and teleworking facilities if it is configured in VLAN Network mode and if a special VPN instance is created by network administrator.

CloudStack has an option for creating VPN, Eucalyptus can create remote desktop, but without establishing VPN. OpenNebula can not create VPN.

3.6 Cryptographic Controls

OpenStack does not use cryptographic controls by default. Using OpenID can increase OpenStack overall security [11]. Neverthless, all OpenStack services does not use any SSL / TLS.

Other analyzed solutions provide different usage of cryptography.

3.7 Security of System Files

OpenStack ensures the security of customers' system files with instance access keys. The customer data is deleted after instance is shut-downed. The data saved in volumes is protected with access keys as well.

Other analyzed open source solutions have similar features.

3.8 Information Security Incident Management

Despite the fact that OpenStack allows cloud service providers to audit logs it provides, it does not possess some internal tool to manage the logs for incident management. However, there are additional tools that provide resource monitoring to help in incident management. Nice tool is Zenoss [25] which manages OpenStack cloud servers, images, instances etc.

All other analyzed open source solutions have the systems for monitoring, both for users and administrators.

3.9 Backup and Disaster Recovery Procedure

OpenStack allows cloud service providers easily to create a snapshot of instances to store the customers data that are in the instances. It has a well designed disaster recovery procedure in case of a disk crash, a network loss, a power cut, etc.

All open source solutions have interfaces to Amazon's AWS providing good solution to backup and disaster recovery.

4 Open source Cloud Solution Security Evaluation

This section evaluates the security for open source cloud solutions described in Section 2 and compared to the OpenStack solution. The basics of security assessment is explained in Section 3 and the evaluation is based on assessment of ISO 27001:2005 [10] control objectives as a guideline for security evaluation.

4.1 Security Evaluation Metric Definition

In this section we define the metrics for security evaluation of the open source cloud solutions compared to OpenStack. Table 1 describes the metrics used in the evaluation. This assessment addressees only relative measures and is not attended to benchmark all solutions, but rather to analyze the OpenStack solution and its security in comparison to the other open source cloud computing solutions.

The suggested metrics include evaluation with 0 if the same level of security is found compared to the OpenStack solution, and correspondingly 1 or -1 if the security level is better or worse than OpenStack.

#	Description
0	The security level is the same as OpenStack
-1	The solution security level is better than OpenStack
+1	The solution security level is worse than OpenStack

Table 1. Metrics for security evaluation in comparison to OpenStack solution

4.2 Security Evaluation

Open source cloud solutions are evaluated in comparison to OpenStack solution according to the metrics presented in Table 1. The evaluation is based on security assessment in Section 3 and assessing each ISO 27001:2005 control objectives that depend of cloud solution.

The results are analyzed by calculation of the average of obtained values for each control objective and the average value of each analyzed open source cloud solution.

Table 2 presents the results of the evaluation. Column ON identifies Open-Nebula, CS identifies CloudStack and Eu identifies Eucalyptus opensource cloud solution.

#	Control Objective	ON	CS	Eu	Avg
10.5	Back-up	0	0	0	0
10.6	Network security management	0	0	0	0
10.7	Media handling	0	0	1	0.33
10.10	Monitoring	0	0	0	0
11.2	User access management	0	1	1	0.67
11.4	Network access control	0	0	0	0
11.5	Operating system access control	0	0	0	0
11.6	Application and information access control	1	1	1	1
11.7	Mobile computing and teleworking	-1	0	-1	-0.67
12.3	Cryptographic controls	1	1	1	1
12.4	Security of system files	0	0	0	0
13.1	Reporting information security events and weaknesses	1	1	1	1
14.1	Information security aspects of business continuity man-	0	0	0	0
	agement				
	average security level evaluation compared to OpenStack	2	4	4	3.33

 Table 2. Security evaluation of open source cloud solutions

4.3 Security Evaluation Analysis

Only 13 out of 39 or exactly 1/3 of the ISO 27001:2005 control objectives depend on cloud solutions. These are presented in Table 2.

Figure 1 depicts the results of our qualitative analysis of the security evaluation. CloudStack and Eucalyptus achieve the best security with 4 security points ahead of OpenStack, meaning that in 4 out of 13 analyzed relevant control objectives (30.77%) they achieve better security. OpenNebula achieves 2 security points, i.e. achieves better security in 2 out of 13 control objectives (15.4%) ahead of OpenStack.

The conclusion is that OpenStack has the worst level of security.

Figure 2 presents another view on the security evaluation comparing open source solutions to OpenStack. It shows the average results of each solution according to the security evaluation compared to the OpenStack security level.

OpenStack achieves equal security level with other open source cloud computing solutions in 7 out of 13 analyzed relevant control objectives (53.85%) and worse security level in 5 relevant control objectives (38.46%).

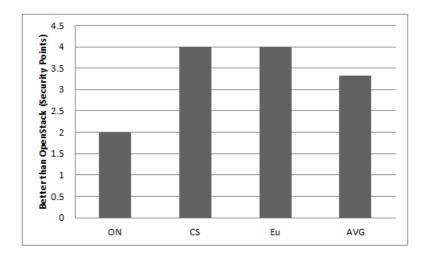


Fig. 1. The qualitative analysis of the security evaluation of other open source cloud solutions compared to OpenStack

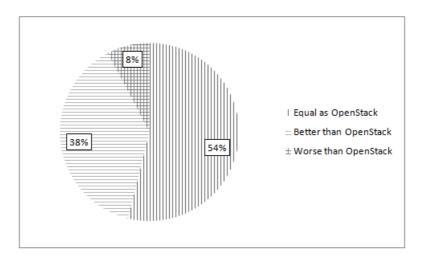


Fig. 2. The average of the security evaluation

Only in 1 out of all 13 analyzed relevant control objectives (7.69%) OpenStack achieves better security level than other open source cloud computing solutions. It concerns Mobile computing and teleworking.

5 OpenStack security pros and cons

This section presents pros and cons of OpenStack security as a result of our security assessment. A comparison with the other open source cloud solutions is presented.

5.1 OpenStack Pros

Positive aspects of security that we found during the security assessment of OpenStack are:

- OpenStack manages user access securely with role based access control;
- OpenStack manages network access securely only if it is deployed in VLAN Network mode;
- OpenStack manages operating system access securely;
- OpenStack ensures information security when using mobile computing and teleworking facilities if it is configured in VLAN Network mode;
- OpenStack ensures the security of customers' system files;
- OpenStack provides log files for incident management but not the tools to process them in real time; and
- OpenStack provides disaster recovery procedure in case of disk crash, network loss a power cut etc.

Comparing with the other open source solutions we can conclude that Open-Stack has better security level than CloudStack and Eucalyptus (same security level as OpenNebula) for mobile computing and teleworking facilities if it is configured in VLAN Network mode because it can configure VPN to the instances of virtual machines. For all the other OpenStack security pros the other open source cloud solutions provide equal or even better security level.

5.2 OpenStack Cons

Despite the advantages and security benefits, the OpenStack provides several security detriments, such as that OpenStack:

- can be configured in Flat Network Manager or Flat DHCP Network Manager where all instances of all projects are assigned IP addresses from one subnet, which does not provide segregation in networks and is incompliant with ISO 27001:2005;
- does not use cryptographic controls for its web services neither for authentication of web GUI for administration - OpenStack Dashboard; and

 does not provide a tool to monitor the system which is necessary for information security incident management.

Comparing the other open source solutions we can conclude that all of them handles better the security issues that are cons for OpenStack. Even more, for the last two cons all other solutions have better security level (Table 2) than OpenStack.

6 Conclusion

IT quality managers have always a dilemma which platform should they select, commercial or open source, on which to drive their services. Both platforms have advantages and detriments. Commercial platforms usually offer more stable, tested, reliable, trusted and less riskiness solutions. However, usually they are more expensive than open source. Open source platforms usually offer acceptable stable solutions. Those solutions are not tested properly and a lot of new versions and sub versions are released in a short period of time. The main advantage of this platform is the small amount of money or even totally free for many products and services. Another huge advantage is the source code that can be used by the customer to redevelop the solution to its requirements.

Both commercial and open source platforms provide solutions to build a private cloud. Almost all commercial and some of the open source provide a solution to build new or interfaces to existing public clouds.

All key commercial cloud providers possess some security certificate as a company. Additionally all of them offer some security services to their customers [18]. Open source solutions provide a small number of security services to the clients or generally do not provide any.

Neither security assessment nor comparative security analysis of the cloud were not performed in the literature so far. In this paper we have analyzed the security issues that OpenStack cloud software possess and what other security tools can be integrated to improve its overall security. The evaluation of the security was realized by assessing the relevant control objectives defined by ISO 27001:2005 and comparing it to the other open source cloud computing solutions.

The results of our assessment show that Eucalyptus and CloudStack have integrated the maximum security level in front of OpenNebula. OpenStack has integrated the least security compared to others solutions.

References

- Amazon: Amazon web services (Mar 2012), http://aws.amazon.com/ what-is-aws/
- Avetisyan, A., Campbell, R., Gupta, I., Heath, M., Ko, S., Ganger, G., Kozuch, M., O'Hallaron, D., Kunze, M., Kwan, T., Lai, K., Lyons, M., Milojicic, D., Lee, H.Y., Soh, Y.C., Ming, N.K., Luke, J.Y., Namgoong, H.: Open cirrus: A global cloud computing testbed. Computer 43(4), 35–43 (april 2010)

- Berger, S., Cáceres, R., Goldman, K., Pendarakis, D., Perez, R., Rao, J.R., Rom, E., Sailer, R., Schildhauer, W., Srinivasan, D., Tal, S., Valdez, E.: Security for the cloud infrastructure: trusted virtual data center implementation. IBM J. Res. Dev. 53(4), 560-571 (Jul 2009), http://dl.acm.org/ citation.cfm?id=1850659.1850665
- 4. Campbell, R., Gupta, I., Heath, M., Ko, S.Y., Kozuch, M., Kunze, M., Kwan, T., Lai, K., Lee, H.Y., Lyons, M., Milojicic, D., O'Hallaron, D., Soh, Y.C.: Open cirrustmcloud computing testbed: federated data centers for open source systems and services research. In: Proceedings of the 2009 conference on Hot topics in cloud computing. HotCloud'09, USENIX Association, Berkeley, CA, USA (2009), http://dl.acm.org/citation.cfm? id=1855533.1855534
- 5. Cirrus, O.: Open cirrus open cloud computing research testbed (Apr 2012), https://opencirrus.org/
- CloudStack: Cloudstack opens source cloud computing (Apr 2012), http: //cloudstack.org
- 7. Eucalyptus: Eucalyptus cloud (Apr 2012), http://www.eucalyptus.com/
- 8. Google: Google app engine (Apr 2012), https://developers.google.com/ appengine/
- 9. Google: Google cloud storage (Apr 2012), https://developers.google. com/storage/
- 10. ISO/IEC: ISO/IEC 27001:2005, Information Security Management Systems - Requirements (2012), \url{http://www.iso.org/iso/iso_catalogue/ catalogue_tc/catalogue_detail.htm?csnumber=42103}
- Khan, R., Ylitalo, J., Ahmed, A.: Openid authentication as a service in openstack. In: Information Assurance and Security (IAS), 2011 7th International Conference on. pp. 372 –377 (dec 2011)
- 12. Microsoft: Microsoft live (Apr 2012), http://www.live.com/
- Microsoft: Microsoft windows azure (Apr 2012), http://www. windowsazure.com/en-us/
- Ng, C.H., Ma, M., Wong, T.Y., Lee, P.P.C., Lui, J.C.S.: Live deduplication storage of virtual machine images in an open-source cloud. In: Proceedings of the 12th ACM/IFIP/USENIX international conference on Middleware. pp. 81–100. Middleware'11, Springer-Verlag, Berlin, Heidelberg (2011), http: //dx.doi.org/10.1007/978-3-642-25821-3_5
- OpenNebula: Opennebula cloud software (Apr 2012), http://Opennebula. org
- 16. OpenStack: Openstack cloud software (Feb 2012), http://openstack.org
- Rajan, S., Jairath, A.: Cloud computing: The fifth generation of computing. In: Communication Systems and Network Technologies (CSNT), 2011 International Conference on. pp. 665–667 (june 2011)
- Ristov, S., Gusev, M., Kostoska, M.: Cloud computing security in business information systems. International Journal of Network Security & Its Applications (IJNSA) 4(2), 75–94 (2012)
- Ristov, S., Gusev, M., Kostoska, M., Kiroski, K.: Business continuity challenges in cloud computing. In: ICT Innovations 2011, Web Proceedings, Skopje, Macedonia (2011)

- Ristov, S., Gusev, M., Kostoska, M.: A new methodology for security evaluation in cloud computing. In: MIPRO, 2012 Proc. of the 35th Int. Convention. pp. 1808–1813 (2012)
- 21. SalesForce: Salesforce sales cloud (Apr 2012), http://www.salesforce.com/
- 22. Tsai, C.L., Lin, U.C.: Information security of cloud computing for enterprises. AISS: Advances in Information Sciences and Service Sciences 3(1), 132-142 (2011), http://www.aicit.org/aiss/global/ppl.html?jname= AISS&yr=2011&vnidx=6
- 23. VMware: Vmware vcloud (Apr 2012), http://www.vmware.com/products/ vcloud/overview.html
- Voras, I., Mihaljevic, B., Orlic, M.: Criteria for evaluation of open source cloud computing solutions. In: Information Technology Interfaces (ITI), Proceedings of the ITI 2011 33rd International Conference on. pp. 137 –142 (june 2011)
- 25. Zenoss: Zenoss service dynamics (Mar 2012), http://www.zenoss.com/

Fleet Management Solution based on Cloud Computing: improving availability of additional services

Dashmir Istrefi¹, Betim Çiço²,

^{1,2} FTI, Polytechnic University of Tirana, Tirana, Albania {D.Istrefi, Betim.Cico}@gmail.com

Abstract. With the help of satellite communication technology today is easy to identify the exact location. Existing solutions support tracking of place (realtime or historical visibility), status of vehicle which mostly are using centralized web server. The purpose of the paper is to describe how solutions in the field of M2M (Machine to Machine) by combining fleet management solution and cloud computing could deliver new technological innovation. The proposed solution uses GPS based technology, GPRS and cloud computing infrastructure. The vehicle used for testing purposes was equipped with the following devices: GPS device with a GPRS module (TELTONIKA FM4200), sensors for identifying fuel level/status, driver identification. All the data that the device sends are transferred to a server running in a cloud system through GPRS traffic. All the data is stored and saved in the cloud and the customer could receive it from any device that can have access to a web page. The solution is reliable, scalable and comprehensive that fits the needs and demands of the existing companies on the market. It scales to all possible fleets of all sizes, and the paper also identifies how the solution can improve operations, asset utilization. Although the number of connected devices, machines that require connectivity is increasing, so far they have not been widely adopted in the South East European region, and the research examines the reason for not fully acceptance of the fleet management solution. The concept was developed taking into consideration some of the most common use cases for companies in different fields that they operate.

Keywords: Fleet Management Solution, Cloud computing, GPS/GPRS technology, Innovative Enterprise Solutions.

1 Introduction

With the development of the telecommunication industries in recent years most of the markets have reached high level of penetration and other developments are needed for further growth of the industry. The number of embedded devices has continuously been increasing in recent years, different kinds of wireless and wired access systems such as GPRS, EDGE, HSDPA, WLAN, WiMax, Zigbee, and Bluetooth enable various kinds of devices such as sensors, actuators and machines to connect into the Internet, which enables novel types of services called as M2M (machine to machine) services. M2M consists of ICT technologies enabling remote measurements and remote control of devices. It includes sending receiving, storing and processing of measured information, and all kinds of actions needed for controlling devices remotely. Thus M2M creates novel value added by connecting large set of machines, vehicles and embedded devices into the M2M networks and service infrastructures and enable remote actions with devices and their services. This enables more real-time control over the company processes, and creates opportunities to increase the service quality [1].

The main contribution of this research is focused to solve the interoperability challenge by providing M2M architecture to enable smooth creation of the M2M service space in automotive transport and logistics space, also referred as Fleet Management. Fleet Management system has brought GPS technology into every type of vehicle used for both tracking and navigation. The tracking part of the device helps the system detect and notify where the vehicle is moving, whereas the navigation part helps the driver to reach the destination. All existing technologies are able to identify and provide this, whereas the customer demand is increasing in order to provide them with additional benefits.

For this research paper TELTONIKA FM1100 and FM4200 [2] Fleet Management terminals are used. The proposed solution is targeting specific use cases where additional sensors are needed in order to provide more security and cost control benefits for the end customer. All these systems are integrated and transfer the data which is maintained in cloud infrastructures. The sensors are used to monitor parameters of the fuel level of the vehicle. This is where, when and how much fuel was filled into the tanks of the vehicles, customized alerts can be set when certain level of fuel level is reached. What the fuel sensor does is that it sends information to the device about the fuel level changes in the tank. Then the device continuously updates the cloud server with the new parameters through the GPRS traffic of the network operator. In cases when the fuel level reaches critical level it then notifies the driver and the owner. Driver identification module enables the driver to start the vehicle engine. This prevents unauthorized usage of the fleet that is owned by the individual/company. In cases when an unauthorized usage is detected the system notifies the owner which vehicle and the location of the vehicle is being used.

The paper is organized as follows. Part 2 of the paper elaborates solutions which are currently available. Part 3 presents the architecture of the Fleet management system model. The last part presents the experimental results of the proposed solution and the conclusions that are drawn.

2 State of art

Various research papers in the field of M2M are made as a proof of concept. Persson in [3] describes how multi-agents are illustrated within a smart parking management application. Other research papers in the field are explained in [4], [5], [6] where the research paper [6] successfully tested and implemented the Alcohol detection in the vehicles that when alcohol will be detected the led from the experiment will glow. Research paper presented in [7] explains how SMS technology can help to locate the vehicle. The proposed work collects positions of the vehicle via GPS receiver and then sends the data of positions to specialized server by the SMS service. After that the position of the mobile vehicle will be displayed on Google Map. The main disadvantage with this solution is that the system offers only live tracking and historical reconstruction.

Paper [8] mainly consists of three steps including vehicle region extraction, Fleet Management and classification. After vehicle detection, a graph-based Fleet Management method is used for building the correspondence between the vehicles detected. And lastly another work which was considered as existing implemented solutions in this paper is described in [9] which is a low cost solution in contrast to expensive solutions. The architecture presented in the paper for Fleet Management system is using wireless sensor technology, which defines the packet structure for communication between the nodes.

Similarly as a substitute technology could be used the communication through satellites but this has many disadvantages and those include the cost of implementing such solutions is ex-pensive for a one time investment but it does not require additional costs for GPRS traffic, neither roaming DATA charges which at the moment are high and are not regulated in the South East European market by regulatory agencies. Another disadvantage is that the device sends parameters to the server after a long period which means that you can't have real-time tracking experience at all.

The system for fleet management offers powerful features, and enabling such a system involves the following components: GPS device for tracking the vehicle, GPRS module on the device to transfer data through cellular networks public infrastructure, cloud infrastructure [10] to host the server and database, and access to GUI (web interface) that the customer has as a point for managing his fleet.

3 Architecture

3.1 Fleet Management System Model

Cellular networks offer a standardized uniform public infrastructure for communication with devices via radio. The upper data speed limit for GPRS of 20-50 kbps is sufficient for transferring data and even higher data rates are attainable with third and fourth generation HSPA/LTE networks.

All vehicles currently are equipped with GPS device that is able to transfer the locations through the signals of the GPS satellites. The user of the system can identify location detail as well as driving history that is related to the route that the vehicle has followed. The GPS device module sends all the parameters using TCP/IP protocol when high reliability is needed sending records more often and modified UDP protocol when the vehicle is mainly out of the home network sending minimum 25 records at once due to saving high DATA roaming charges.

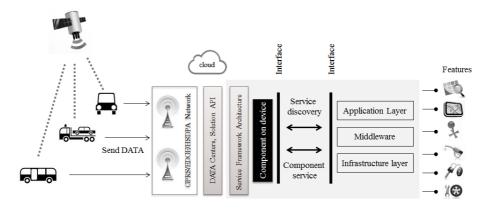


Fig. 1. Overview of Fleet Management system including services framework architecture in the research paper.

Service framework provides generic functionalities for application development; the purpose is to manage components on the device. For example in the experimental phase we have constructed service components for 1-wire devices, fuel tank sensors, impulse counters, alarm buttons, door sensors, ignition etc. Interfaces in the figure are illustrated showing both service discovery and specific protocols.

According to this solution important findings addressed in this paper include:

- Fuel consumption detection
- Driver identification

Fuel consumption can be delivered in some of the following ways. Fuel flow meters [11] are suitable for measuring low viscous liquids like gasoline, diesel, water, beverages, or juices. In principal for safety and accuracy issues this option is not appropriate for the solution because for small amount of fuel consumption flow meters are not accurate. Another way is to use fuel tank level sensors that exist in most of the cars which shows the approximate fuel level in the driver's indicator panel. This solution is suitable only for big trucks but even there are limitations for trucks that have cylindrical form of tank because additional software adjustments have to be made that in the end reduces accuracy. Third way is to connect to vehicles CAN-bus (controller area network) where SAE J1939 is the vehicle bus standard used for communication and diagnostics among vehicle components. Standardizes parameters are available and among them is the fuel consumption as well. Limitations of this method are that it that limited brands are supported and it's difficult to connect. Apart from that all three methods share common drawbacks of the information accuracy or damage that may be done to the vehicle or the GPS module while integrating it.

Driver identification module takes into consideration delivering communication over the serial protocol providing combinations of memory, mixed signal, and secure authentication functions via single contact serial interface. Module communicates via 1-Wire protocol that is explained in [12].

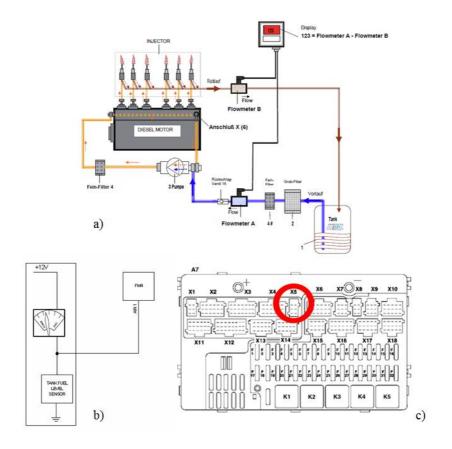


Fig. 2. a) Bio-Tech flow meter [11], b) Taltonika fuel tank level sensor [2], c) X5 connector located in the fuse box for CAN wires.

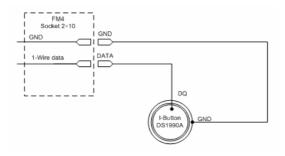


Fig. 3. Dallas key iButton DS1990A connection scheme.

Value added service concept requires communication between the device and other devices like alarm, fuel, door sensors. There are no set of communication standards

for this type of applications, both fixed and wireless could be used for this purpose. External communication interfaces can support industry standards like ZigBee IEEE 802.15.4 and enable the creation of an open system infrastructure where additional devices such as sensors, actuators and machines could be connected over the time to the Fleet Management service framework. It's a low power consumption, low cost, wireless networking protocol targeted for automation and remote control [13].



Fig. 4. ZigBee network [14], a ZigBee Device Types, ZB Coordinator – 15.4 PAN Coordinator, ZB Router 15.4 FDD/Coordinator, ZB End device 15.4 RFD.

Experimental part of the research proposes different solution which is more feasible for implementation than the abovementioned fuel and driver identification techniques for delivering the solutions.

3.2 Cloud infrastructure

Utilization of the server spaces is high as there are not many using cloud computing infrastructure and web portal for retrieving data. The proposed solution gathers information and transmits data to the server stored in cloud infrastructure. By making data available in the cloud, it can be more easily and ubiquitously accessed, often at much lower cost, and even gain more flexibility by deploying a full service set without having to set up base infrastructure to support. In terms of scalability cloud computing allows the service to scale over a wide demand range [15].

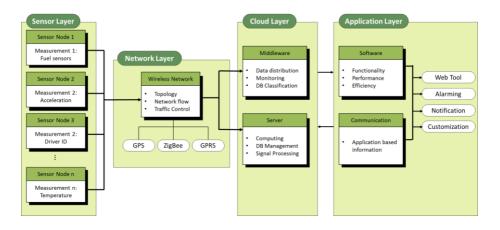


Fig. 5. Diagram with information flow and how all layers that are related to one another and displayed on a single graphical interface to the customer.

Sensors range from different parameters such as acceleration, measurement of direction, driver identification, fuel level, temperature, alcohol detection, all of which through wireless network technologies transfer the parameters to middleware and server hosted as components of the cloud layer.

The emergence of distributed cloud-based applications and services brings plenty of opportunities to improve cost, scale, reliability and performance. This represents a fundamental yet massive shift in "computing". In other words, it is the next natural step in the evolution of computing in general, and IT services delivery in particular. The key idea of this paradigm is to provide service, into which a customer could plugin instantly and modify the service to access centrally located services.

The application layer enables customers have access to web tool, set up alarming and notification messages to different devices ranging from Smartphones, to any other device that has access to a web browser.

Application of Fleet Management to Clouds comprises of:

- Easy purchase of additional licenses for every vehicle through the web tool
- Branding/Customization of the GUI according to every customer need
- Alarming and notification to device by your choice
- Enhancing the model to support bring your own device model and not relying fully to one model as is the case with existing solutions, subject of a development in some other stage of research

The term "Cloud Computing" [10] refers to an online delivery and consumption model for business and customer services, ranging from IT services like Software-as-a-Service (SaaS), and share the fundamental approach of running server components elsewhere, over the Internet.

Extensible cloud architecture is displayed in the following graph.

End user GUI						
\$	\$	\$				
Management Interfaces						
Administrator management	Vehicle information	Customer branding				
DATA Aggregation	layer					
GPS device Interfaces						
DATA Aggregation layer						
Data Centers						
DATA Center	DATA Center	DATA Center				

Fig. 6. Extensible architecture.

- GPS device interfaces: these interfaces do metering based on how much is consumed and alerts predefined by the customers in order to alert them about their fuel consumption, driver identification, geo-fencing.
- Communication layer: because of diverse metering interface systems in the open market, there is a need for a common layer to interface with different metering and reporting systems.
- Management Interfaces: these interfaces control the business management functions such as customer branding of the portal, vehicle information such as registration, servicing, administration of users/privileges through the DATA aggregation layer.
- DATA aggregation layer: This is a common data collection repository designed for frequent updates. This is the collection point for all data across data centers with other connected agencies.

4 Experimental Results

Driver identification and fuel module were tested and implemented for the purpose of the research paper. Compared to the abovementioned options the following experiment was conducted. For driver identification proximity reader MF7 featuring medium range and small dimensions was used. Interface uses 9.6K Baud Serial ASCII (RS232) [16]. The output format is customized and programmed that it allows apart from driver identification to distribute club cards for passenger identification as well. Fuel module is completely different from the above available options as it uses calculated consumption method which is software developed and provides as good or better accuracy with cheaper implementation. It is based on 3 main parameters for

calculating fuel consumption and those include engine hours in stop, driving in a city consumption per mileage and open road consumption.

The results show information relevant for the customer and receive real-time alerts in order to maintain control over the vehicle fleet.

Separate main modules as shown are shown as a result of work of this paper:

- Tracking (live and historical),
- Fuel (cost control feature),
- Fleet (maintain all relevant information for the vehicles like insurance expiry, oil/filter change), and
- Users (administration of privileges and different levels for using the solution)

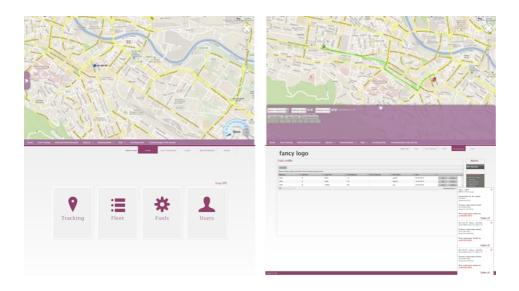


Fig. 7. Results from the implemented solution a) live tracking, b) historical reconstruction, c) administration panel, d) fuel administration and servicing alarms.

5 Conclusions and Future work

The paper is a basis for future improvement of the solution providing an integrated support for the development of the application.

As future work planned is to improve the existing system is that through forms of interviews with different companies that their mainly focus of the business is in the transport and logistics following daily tasks to identify problems which are currently not solved by any kind of technology which they are using:

- Reports or alerts when vehicles are used out of working hours for private purposes
- Places that are visited outside of the defined zone that is allowed

- When vehicle is moving with passengers on board but the taximeter is not turned on
- Block the motor of vehicle in case of unauthorized usage.

Additionally improvement on the Service Framework Architecture to support additional M2M verticals apart from Fleet Management is planned for future research.

References

- J. Latvakoski, T. Hautakoski, T. Vaisanen, J. Toivonen, A. Lappalainen, T. Aarnipuro, Secure M2M Service Space in Residential Home, COMSWARE '09, Dublin, Ireland, 2009
- 2. Teltonika, Official web site (online: http://www.teltonika.lt/en/pages/view/?id=10), accessed 25 April, 2012
- 3. C. Persson, G. Picard, F. Ramparany, A Multi-Agent Organization for the Governance of Machine-to-Machine Systems, International Conference on Web Intelligence and Intelligent Agent Technology, 2011
- 4. P. B. Jeon, et al., Semantic Negotiation-based Service Framework in an M2M Environment, International Conference on Web Intelligence and Intelligent Agent Technology, 2011
- 5. G. Coulson, et al., Flexible experimentation in Wireless Sensor Networks, Communications of the ACM, January 2012
- 6. A. Alexe, R. Ezhilarasie, Cloud Computing Based Fleet Management Information Systems, IJCST, Vol. 2, Issue 1, March 2011
- 7. Thuong Le-Tien, Vu Phung, Routing and Tracking System for Mobile Vehicles in Large Area, The Dept. of Electrical Electronics Engineering, HCM University of Technology, Vietnam
- 8. Jin-Cyuan Lai, Shih-Shinh Huang, Chien-Cheng Tseng, Image-Based Fleet Management and Classification on the Highway, Dept. of Computer and Communication Engineering, National Kaohsiung First University of Science and Technology
- 9. Aravind, K.G.; Chakravarty, T.; Chandra, M.G.; Balamuralidhar, P., "On the architecture of Fleet Management system using wireless sensor devices", TCS Innovation Labs., Tata Consultancy Services, Bangalore, India
- 10. K. Stanoevska-Slabeva, et al., Grid and Cloud Computing: A Business Perspective on Technology and Applications, Springer, 2010
- 11. BIO-TECH e.K., Flow meter product range, (online: http://www.btflowmeter.com/en/products.html), accessed 15 May, 2012
- 12. Dallas Semiconductor/Maxim, 1-Wire Protocol, (online: http://coecsl.ece.illinois.edu/ge423/sensorprojects/1-wire_full.doc), accessed 15 May, 2012
- S. E. Ergen, ZigBee/IEEE 802.15.4 Summary, September 2004, (online: http://www.prism.uvsq.fr/~mogue/SENSORS/Sensor%20%20Net/MAC%20pro/zigbee_8 02.15.4.pdf) accessed 22 April, 2012
- 14. M.J. Skibniewski, W. Jang, Ubiquitous computing: object tracking and monitoring in construction processes utilizing ZigBee networks
- 15. ACM queue, Cloud Computing: An Overview, 2009
- Ivar a.s., Proximity Reader MF7, (online: http://www.ivar.cz/katalog1/MF7_20/MF7-20_technicky_manual_TM951060-02.pdf), accessed 1 May, 2012

Load balancing in distributed systems and ad hoc networks using a fairness-oriented approach

Stylianos Dimitriou^{1,*}, and Elias Christodoulou¹,

¹ Technological Educational Institute of Serres, Serres, Greece, {sdimitr, christodoulou.ilias}@gmail.com

Abstract. The widespread use of portable mobile devices with networking capabilities and the necessity to interconnect those using low-cost ad hoc networks, imposes a new perspective over traditional wired networking. Apart from low connectivity and variable delay issues, protocols for mobile ad hoc networks should also take account of the energy constraints evoked by the use of batteries. Routing becomes a major concern in such networks, since communicating nodes often become routing nodes as well and their lifespan is decreased by relaying packets from other nodes. In order to address this issue, we propose a generalized load balancing algorithm, namely Fairness-oriented Load Balancing. Although our proposal is not exclusive for networking, it can be easily applied in mobile networks whose performance depends heavily on the energy-efficiency of the underlying protocols. Fairness-oriented Load Balancing can be adapted to cover the needs of a wide range of applications and systems. The initial results indicate that our proposal outperforms traditional process assignment in terms of system fairness while maintaining a stable behavior.

Keywords: Wireless networks, MANET, DTN, routing, fairness.

1 Introduction

Fairness in computer networks is a term that can have different interpretations depending on the type of network, the protocol layer or the application utilized. Most times, we consider fairness as a uniform distribution of network resources among users with similar needs [2]. In other cases, we consider fairness as a uniform distribution of the work load of a network. Regardless of the definition, users often value high system fairness more than high performance, as it ensures that all nodes will enjoy similar benefits and undertake similar amount of burden. In order to address the problem of occasionally uneven load distribution among nodes of a computer network, we propose a novel approach, namely Fairness-oriented Load Balancing (FLB). FLB is a generalized algorithm that aims to distribute fairly the work load of a distributed system to a number of processors. If we consider a mobile

^{*} Corresponding author: e-mail: sdimitr@gmail.com, address: Technological Educational Institute of Serres, end of Magnesia street, 62124, Serres, Greece.

ad hoc network (MANET) as a distributed system where nodes (processors) exchange packets (work load) among themselves, then we can utilize FLB as a routing mechanism focused on increasing system fairness. Packets will not be routed exclusively through a limited group of nodes and, thus, the lifespan of batterypowered nodes will not be decreased disproportionately. Even though load balancing in computer networks is not a new problem, it is often regarded as an issue of performance increase, rather than fairness. The fairness-oriented perspective we adopted, is imposed by the advances that took place in mobile networking during recent years.

The recent scientific interest on space and mobile communications in general, has created a new research area for computer network engineers. Whereas in the past decades communication among space nodes was only possible with manual data transmission in predetermined time instances, in the recent years scientists are researching ways to minimize the human interference required and increase the autonomy of the communication process. Indeed, the necessity to exploit every communication opportunity that might occur in space conditions, as well as the need to transmit a big portion of earth communications via space, has shifted the research interest towards new space protocols. Although these protocols are required to operate in significantly different conditions to Earth, they are designed to behave similarly to Internet protocols in terms of connectivity, autonomy and reliability. At the same time, there are several types of terrestrial networks which manifest similar characteristics to space networks, such as vehicular networks and MANETs. Current implementations have several limitations and may not scale properly in big networks. Thus, when it comes to interconnecting mobile devices with long and variable delays, the need for a more flexible approach is apparent.

In mobile wireless networks, the network infrastructure is created ad hoc. All the communicating devices are simultaneously sending, routing and receiving battery-powered nodes. This approach is problematic when it comes to routing packets: since nodes are forced to transmit not only their own packets but also incoming packets from other nodes, their lifespan can be decreased significantly. Too many forwarded packets can jeopardize the smooth operation of the entire network. Fairness-oriented Load Balancing is an algorithm that can be used in order alleviate this issue. It increases fairness among mobile nodes by ensuring that packets are forwarded by all the nodes, and not by a limited group. To achieve its aim, FLB develops the routing strategy based on the properties of the packets the nodes have previously forwarded. These properties can refer to the packet (packet size, overhead) or the node (distance, available battery).

In order to take account of all these properties simultaneously, FLB is designed as a multi-criteria algorithm that has the ability to review the available choices and decide quickly the best strategy. FLB is a generalized algorithm that can be applied to a variety of distributed systems from computer networks to parallel computational systems and HR management, without affecting its core functionality. The algorithm uses two main concepts: process and processing units or units. Each process is characterized by its properties and each unit by its history. It is up to a central authority to assign the processes to the units in a way that, in the end, all units have been assigned processes with similar properties. Some of FLB main characteristics are:

- Scalability. The algorithm can perform regardless of the size of the system (number of units, number of process properties utilized etc).
- Need for a central authority, which is responsible to assign processes to units. Depending on the system, it can be a sending node, a central processor or a human manager.
- Discrete time. We divide the time in discrete slots or steps. Processes are assigned in the beginning of each time slot and are processed during the time period of this slot. All units are available at the beginning of the next time slot. Nevertheless, as we show in section 3, this assumption does not limit us only in clock-synchronized systems.
- Memory. Units have recorded history, where they register what type of processes they have previously processed. Future assignments take into account the history of each unit.

Following, we analyze FLB in detail, along with its impact on system fairness. We also evaluate it through some initial simulations and compare it with serial process assignment. The rest of the paper is organized as follows. Section 2 discusses the related work in load balancing on computer networks. Section 3 describes the generalized FLB algorithm. We continue with FLB evaluation in section 4, where we perform a set of simulations that aim to support our proposal. In section 5, we explain how we can apply FLB to delay tolerant and mobile networks in general, and in section 6 we conclude and set the framework for future work.

2 Related Work

Load balancing in computer networks usually refers to the distribution of a web service load through multiple servers [1]. Such services operate on the Application Layer (eg HTTP, FTP, DNS) and depend heavily on the fluctuations of traffic patterns. In these cases, demand occasionally exceeds supply and processors tend to operate always at full capacity. Resources are a precious commodity which we should fully exploit and fairness is important as long as the primary requirement for high performance is satisfied.

Moving to lower protocol layers, load balancing issues arise in routing and active queue management. Routing protocols implement node and path discovery through various algorithms and are responsible for traffic distribution in the network. When specific areas become congested, packets should be rerouted to less occupied network channels. Congestion events in wired networks are usually transient and, apart from node synchronization issues, have no serious side-effect for the network. Nevertheless, routing in mobile ad hoc networks is a little more intricate [3] due to their dynamic behavior. The network topology changes constantly, bandwidth varies proportionately to node proximity and energy limitations become progressively restrictive. In [11] we have a categorization of MANET routing algorithms and in [10] a short comparison. Although some of the challenges mentioned for MANET routing are power aware routing and network overhead, there are few proposals

working in this direction [6]. Minimum Total Transmission Power Routing (MTRP) calculates the total transmission power for all available routes and picks the one with minimum total transmission power. Its main disadvantage is that it only considers the energy constraints of entire routes, rather than individual nodes. Minimum Battery Cost Routing (MBCR) works in a similar way, by summing the available battery of all nodes in a single route, while having the same drawbacks as MTRP. Addressing this problem, Min-Max Battery Cost Routing (MMBCR) takes account the least-powered node, without ensuring that the best route is eventually chosen. Finally, Conditional Max-Min Battery Capacity Routing (CMMBCR) takes account the least-powered nodes per route. What these proposals have in common is that they consider power-aware routing as a battery-exclusive issue, ignoring other parameters that might arise.

Other works [9] [13] that propose metrics and methodologies for implementing power-aware routing, mention more parameters but in a separate way, and do not follow a unified approach. In [4] and [8] the authors propose routing solutions for MANETs which are inspired by the Modified Ant Colony Optimization framework. Although the authors make some interesting points, the proposed solutions are mainly focused on resource exploitation than fairness. In a similar perspective, the authors in [5] propose a Position-Based Hybrid Routing (PHBR) algorithm which aims more on effective use of bandwidth by reducing the routing overload, rather than raising power-aware issues.

Load balancing is also the subject of some works which focus on Active Queue Management. Paper [7] proposes Delay Tolerant Queue Management (DTQM), a queue management solution for deep-space networks. DTQM uses two types of buffers (Connectivity and Non-Connectivity) and implements load balancing between the two, in order to achieve increased node lifespan. A similar priority based algorithm is proposed in [12]. Although it also focuses on wireless networks, it only addresses dropped packets, rather than unnecessary retransmissions.

Of course, load balancing issues exist in other distributed systems as well, such as parallel computing and human resources management. Depending on the application, we may be interested in high performance, high yield, low execution time or low latency. Group management is the most interesting application, as worker satisfaction is a multidimensional problem and several parameters have to be optimized in order to achieve an optimal result.

3 Fairness-oriented Load Balancing

We will now proceed with a detailed description of the FLB algorithm. At this point, we will not attempt to restrict our description to a specific application (eg computer networks or parallel computing systems). This means that our analysis will be generic, instead of networking-oriented. Our aim is twofold; show that our proposal can promote fairness in any distributed system and investigate the potential benefits for computer networks. Throughout our description we will use two basic notions: the process and the (processing) unit. A process is assigned to a unit by a central authority and the unit processes the process in a specific time period.

602

We consider a distributed computing system which consists of **M** units and has to processes **N** processes in a specific period of time. These processes can be generated a priori or on-the-fly. We assume that these processes are not identical to each other, but that they have different characteristics, such as processing time or complexity. Our objective is to assign the processes in a way that, in the end, all units will have served processes with similar properties. These properties are depicted by a set of parameters ($\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_k$) whose values can be determined by the central authority or by the inherent characteristics of the process. We consider discrete time and divide it in time steps. At the beginning of each step, **n** processes are available for assignment. These processes should be distributed to $\mathbf{n} \leq \mathbf{M}$ units, while **M-n** units will be inactive. During the time period of one step, the processes are being processed. After **N/n** steps, all the processes have been processed by the units. Although it facilitates analysis, it is not necessary to have a constant **n** or **M** for each round. That is, processes can be generated with stochastic patterns and units can become available or unavailable at specific times.

Processes are assigned to units based on their history. History is determined by the processes the unit has previously processed. Similarly to process properties, each unit's history is depicted by a set of parameters $(\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k)$ which are equal in number to the process properties. In order to choose the most appropriate unit to assign the next process, we calculate the Euclidean distance between the process and the units (assuming a k-dimensional Cartesian coordinate system, Equation (1)) and we pick the unit with the biggest distance.

$$D = \sqrt{\sum_{i=1}^{k} (u_i - p_i)^2}$$
(1)

After the process assignment, the history of the unit is calculated as follows (Equation (2)).

$$u_i = 0.9 \cdot u_i + 0.1 p_i, \ i \in [0, k]$$
⁽²⁾

The choice to pick the unit with the biggest distance is justified if we consider that a big distance between process properties and unit history means that the unit has previously processed processes very different from the current process. For example, assume we have a system where processes have only one property (completion time) and three units U₁(30), U₂(8), U₃(5). The history of each unit denotes that U₁ has previously processed processes with big completion times, whereas U₂ and U₃ have processed processes with small completion times. A process P₁(4) is generated. It is apparent that if we would like to promote fairness, we should assign P₁ to U₁. The history of U₁ will become 27.4 (0.9·30+0.1·4=27+0.4=27.4).

Some more notes on the algorithm:

- Before the system operation, the history of all units is set zero. We determine the first assignment randomly. Then, we set the history of this unit equal to the properties of the first process. We continue similarly, until we assign one process to each unit.
- In order to avoid always assigning processes to the same unit, we can impose a 10% restriction rule. According to this rule, no unit can be assigned more

than 10% the number of processes than another unit. However, this restriction assumes a static system, which is not always the case in real-life systems and might have a significant impact on computation overhead.

• The number of units N can fluctuate during the system operation. Distance measurements can be done regardless of these fluctuations.

To further explain FLB functionality, we will use a simple example and assume a distributed system of 2 units. The system will operate for 2 time slots and will execute 2 processes per time slot (4 processes in total). Each process will have 2 properties. We represent the units with their histories as $U_1(0, 0)$ and $U_2(0, 0)$ and the processes with their properties as $P_1(10, 0.5)$, $P_2(3, 3)$, $P_3(8, 1)$ and $P_4(7, 4)$. We will explain the system operation step by step.

Phase	Processes	U ₁ history	U ₁ distance	U ₂ history	U ₂ distance	Steps
Beginning	$P_1(10, 0.5)$	(0, 0)	-	(0, 0)	-	Α
	$P_2(3, 3)$					
	$P_3(8, 1)$					
	$P_4(7, 4)$					
First time	$P_1(10, 0.5)$	(0, 0)	10.01	(0, 0)	10.01	В
slot	\rightarrow					
	$U_1(0, 0)$					
$P_1(10, 0.5)$	$P_2(3, 3)$	(10, 0.5)	7.43	(0, 0)	4.24	С
$P_2(3, 3)$	\rightarrow					
	$U_2(0, 0)$					
Second	$P_3(8, 1)$	(10, 0.5)	2.06	(3, 3)	5.38	D
time slot	\rightarrow					
	$U_2(3, 3)$					
$P_3(8, 1)$	$P_4(7, 4)$	(10, 0.5)	4.61	(3.5, 2.8)	3.7	Ε
$P_4(7, 4)$	\rightarrow					
	$U_1(10, 0.5)$					
End		(9.7, 0.85)		(3.5, 2.8)		F

Table 1. Example of a simple distributed system.

(A) In the beginning, we set the histories of all units to zero. Our system consists of two units and has to process four processes in total.

(B) The first time slot consists of two steps. During the first step, only two processes can be processed; P_1 and P_2 . Since both of the units have zero history, we assign P_1 to U_1 randomly and set the history of U_1 equal to the properties of P_1 . U_1 becomes unavailable for this time slot.

(C) During the second step, we assign P_2 to U_2 . We update the history of U_2 accordingly.

(D) During the second time slot, two more processes can be processed; P_3 and P_4 . We calculate the distances from P_3 to U_1 and U_2 (Fig. 1). The distance from P_3 to U_2 is bigger than from P_3 to U_1 . Thus, we assign P_3 to U_2 and update its history. U_2 becomes unavailable.

(E) Since U_2 is unavailable, we assign automatically P_4 to U_1 and we update its history.

(F) There are no more processes to process. The final histories for each unit are $U_1(9.7, 0.85)$ and $U_2(3.5, 2.8)$.

We evaluate FLB by comparing it with serial/FIFO process assignment where processes are assigned as they are generated to the first available unit. As an evaluation metric, we use the average values of the assigned processes for each unit. In FLB, we have $U_1 \rightarrow (8.5, 2.25)$ and $U_2 \rightarrow (5.5, 2)$. If we had opted for FIFO assignment, we would have assigned P₁ and P₃ to U₁ and P₂ and P₄ to U₂ and the corresponding average values would have been $U_1 \rightarrow (9, 0.75)$ and $U_2 \rightarrow (5, 3.5)$. Although the yield in both cases is the same (all processes are processed in two time slots), FLB outperforms FIFO in means of system fairness. The differences might seem trivial at this example. Nevertheless, as we will see in section 4, in bigger scale systems the impact of FLB is more apparent.

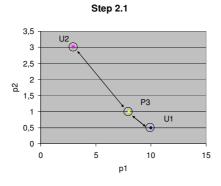


Fig. 1. Measurement of distances from units U_1 , U_2 to the process P_3 .

Before we conclude our description, we should raise the restrictions we set before concerning the discrete operation time and the constant process duration. Continuous time is emulated by assuming only one process assignment per step. Thus every assignment is independent of the previous. Variable process duration is emulated by assuming variable unit availability: we consider a unit unavailable for the next assignments, until it has processed its assigned process.

4 Experimental Validation

In order to evaluate FLB experimentally, we have implemented the core algorithm in C++. Using this implementation, we perform three experiments with varying unit and process number. We consider constant unit number throughout each experiment and three process properties: Duration, Complexity and Resources. Before each experiment we define the total number of units available, the total number of steps and the properties of each process. The values of each process characteristic are picked randomly. To increase the validity of our experiments, we will perform each experiment twice, each time with different process order. The number of processes that can be processed per step are always one.

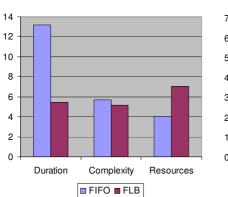
We compare FLB with FIFO process assignment. Since we are interested mainly in fairness, we will compare the deviations of each property from the average value. Less deviation means that units are treated more equally by the distributed system.

4.1 First Experiment

For the first experiment, we consider 3 units and 15 processes in total (Table 2). This is a simple scenario with low complexity. The values of the properties of each process are chosen randomly.

process ID	Duration	Complexity	Resources
proc1	0.68	8.46	8.84
proc2	3.20	8.08	4.58
proc3	8.32	5.22	6.54
proc4	5.01	6.64	1.34
proc5	4.49	0.52	3.52
proc6	9.66	16.24	4.75
proc7	4.01	2.76	17.43
proc8	8.11	23.67	19.50
proc9	11.46	8.22	19.76
proc10	5.99	15.29	3.43
proc11	64.36	56.93	36.82
proc12	36.65	4.10	83.16
proc13	12.59	72.8	88.86
proc14	98.25	14.80	24.25
proc15	89.03	11.02	23.57

Table 2. Process properties for the first experiment.



First Experiment (a)

First Experiment (b)

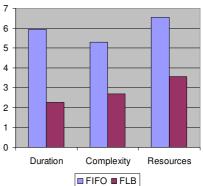


Fig. 2. First experiment deviation results.

606

We will perform the experiment twice, with different processes order each time. We will measure the deviations of each unit average values from the total average. The results from each execution can be seen in Fig. 2.

Our first remark is that FLB succeeds at decreasing the deviation of unit histories, almost for the total of process properties. This decrease means that the three units were assigned processes with similar properties. That is, the aggregate values for all the units, concerning Duration, Complexity and Resources, tend to be the same for all the peers. This results in fairness increase, which can be perceived easily by users.

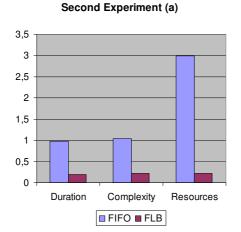
Another thing we notice is that when we shuffle the processes, the results may fluctuate significantly. This behavior is more apparent in FIFO assignment, where processes are assigned based only by their precedence. FLB tends to have a more stable behavior, since it assigns processes dynamically to the units.

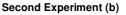
4.2 Second Experiment

For the second experiment, we consider a bigger-scale system which consists of 2 units and has to process 200 processes. Once again, we choose the process properties randomly (Table 3); however the ranges this time are bigger.

Table 3.	Process properties	for the second	experiment.
----------	--------------------	----------------	-------------

process ID	Duration	Complexity	Resources
proc1 – proc50	rand(0,1)	rand(0,1)	rand(0,1)
proc51 – proc100	rand(0,25)	rand(0,25)	rand(0,25)
proc101 - proc150	rand(0,50)	rand(0,50)	rand(0,50)
proc151 – proc200	rand(0,100)	rand(0,100)	rand(0,100)





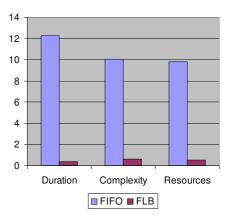
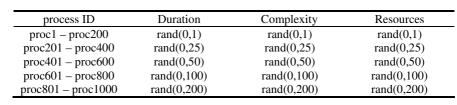


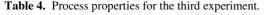
Fig. 3. Second experiment deviation results.

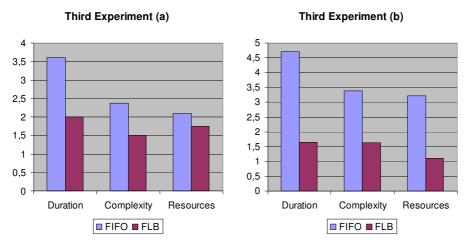
The differences between FIFO and FLB in this scenario are more apparent (Fig. 3). Since the units are now two, it is easy for FIFO to randomly assign processes to the least suitable unit, thus promoting unfairness. On the other hand, FLB manages to achieve high-quality results for the same reasons: since the units are only two, it is very easy to monitor the history of each one and assign the appropriate process. Although the results depend heavily on the values of the properties we picked, they still indicate FLB superiority over FIFO assignment.

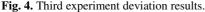
4.3 Third Experiment

For the third experiment, we consider a more complex scenario with 10 units and 1000 processes (Table 4).









In this last experiment differences are not as notable as before (Fig. 4), due to the increased number of units. Once again, this is due to the unpredictable behavior of FIFO assignment, which may randomly promote fairness when several units exist. Nevertheless, even in under these conditions, the decrease on histories deviation spans from 20% to 60% and FLB manifests a stable behavior.

5 Applying FLB to Mobile Networks

The experiments we conducted in section 4 consist an initial validation of our algorithm's efficiency. Although in no way exhaustive, they depict FLB's tendency to promote fairness in a variety of conditions, while maintaining a stable behavior regardless of the values of processes properties. Since we attempted to define our algorithm as abstract as possible without limiting ourselves to a specific type of distributed system, our analysis is valid for any distributed system. This means that we are free to restrict FLB to a specific application without loss of generality.

Several applications in computer networks can be solved if we consider the network as a distributed system, where traffic represents the work load and nodes represent the processors. In this section, we will deal with packet forwarding and routing in Delay Tolerant Networks and mobile networks in general. Similar things apply in multi-queue management. If we try to correspond the terms we used in Fairness-oriented Load Balancing to mobile network routing, then we correspond packets to processes and routing nodes to units. Packet properties that we should take into account can be packet size, expiration time, transmission delay etc. The steps that we will follow in order to determine the ideal path for a packet through the mobile network, are similar to process assignment in distributed systems.

A sending node has to send a packet. Through path and node discovery a number of possible routes appear. The sending node requests the history of each of the potential nodes that can forward the packet. Based on these data and on the packet properties, it selects the forwarding node. This forwarding node receives the packet and proceeds to update its history accordingly. During the network operation, nodes can enter and exit the network, several packets can be transmitted and new routes can be generated without affecting the algorithm efficiency. If we manage to achieve a state where all the nodes forward packets with similar characteristics, we will ensure an increased network lifespan, which in turn will promote performance.

6 Conclusion and Future Work

We have proposed a new load balancing algorithm which can be used to promote energy efficiency in mobile networks. FLB increases fairness by efficiently assigning processes to processing units in a distributed system, in a way that the aggregate loads of all units are equal. Using simulations we showed how FLB behaves and compared it with traditional FIFO assignment. FLB outperforms FIFO assignment in terms of fairness and stability. It can scale properly and can be adjusted to a number of conditions and parameters.

Even though the initial results are encouraging, there are still several issues we should investigate further. FLB dictates that units inform the central authority about their history data. In computer networks, this message exchange might increase the transmission delay and affect significantly the communication process. Continuous message exchange might, in fact, consume more energy than before and should be avoided. Moreover, although in several distributed systems a central authority can monitor process assignment, this is not always the case. Mobile networks with several

scattered nodes can suffer from misbehaving nodes who will attempt to gain more than their fair share by modifying their history data. Additionally, we could review the node-oriented perspective for routing applications and instead use a path-oriented approach. Finally, evaluating the algorithm in a network simulator or using a testbed, will assist in calibrating the mechanism to actual conditions and augment its potential for real-life development.

References

- 1. Bourke, T.: Server Load Balancing. O'Reilly Media (2001)
- 2. Chiu, D.-M., Jain, R.: Analysis of the increase and decrease algorithms for congestion avoidance in computer networks. Computer Networks, ISDN Syst., 17(1):1-14 (1989)
- 3. Corson, S., Macker, J.: Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations. RFC 2501 (1999)
- 4. Iyengar, S., Pattnaik, S.: Solving the MANET Routing Problem using Ant Colony Algorithm, Bachelor of Technology thesis (2010)
- 5. Kara, R., Ozcelik, I., Ekiz, H.: A new routing algorithm in MANETs: Position based hybrid routing. Scientific Research and Essays, vol. 5, no. 3, pp. 328-338 (2010)
- Kim, D., Garcia-Luna-Aceves, J.J., Obraczka, K., Cano, J.-C., Manzoni, P.: Routing Mechanisms for Mobile Ad Hoc Networks based on the Energy Drain Rate. IEEE Transactions on Mobile Computing, vol. 2, no. 2, pp. 161-173, (2003)
- Lenas, S. A., Dimitrou, S. Tsapeli, F., Tsaoussidis, V.: A Queue Management Architecture for Efficient Delay Tolerant Networking. International conference on Wired/Wireless Internet Communications (2011)
- Mamoun, M. H.: A New Proactive Routing Algorithm for MANET. International Journal of Academic Research, vol. 2, no. 2, pp. 119-204 (2010)
- Rishiwal, V., Yadav, M., Verma, S.: Power Aware Routing in Ad Hoc Wireless Networks. Journal of Computer Science and Technology, vol. 9, no. 2, pp. 102-109 (2009)
- 10.Talooki, V.N., Ziarati, K.: Performance Comparison of Routing Protocols for Mobile Ad hoc Networks. Asia-Pacific Conference on Communications (2006)
- 11. Taneja, S., Kush, A.: A Survey of Routing Protocols in Mobile Ad Hoc Networks. International Journal of Innovation, Management and Technology, vol. 1, no. 3, pp. 279-285 (2010)
- 12.Toklu, S., Simsek, M., Yildiz, O., Bay, I., Simsek, A., Ackayol, M. A.: A Priority Based Protocol And Load Balancing For Queue Management on Wireless Networks. ICWN (2008)
- 13.Woo, M., Singh, S., Raghavendra, C. S.: Power-Aware Routing in Mobile Ad Hoc Networks. ACM/IEEE MOBICOM (1998)

A New Public Key Cryptosystem with Key Escrow Capabilities

Ruxandra Olimid

University of Bucharest, Bucharest, Romania ruxandra.olimid@fmi.unibuc.ro

Abstract. Public key cryptosystems with key escrow capabilities allow a trusted third party to benefit of the same abilities as the owner of a private key. We propose such a cryptosystem that is inspired by the recently introduced notion of ASGKA (Asymmetric Group Key Agreement) and analyze its security and its performance. Because the real case scenarios impose a hierarchical organizational structure with multiple levels (the organizational chart of a company for example), we extend our system to allow this feature.

Keywords: key escrow, asymmetric multi-key cryptography, group oriented cryptography

1 Introduction

Let's consider the case of a company nowadays. The communication inside the company or with outsiders (clients or partners) should satisfy at least some basic security requirements that include privacy (encrypted communication). But what if an employee cannot fulfill his duties (he could be in holiday or he could have restricted access to the company network)? For example, he could not access an important e-mail. An obvious solution is that the manager should be able to fulfill the employee's duties.

Therefore, the manager should have some privileged rights. He should be able to supervise the activity of any of his subordinates, by decrypting and reading their messages, while they remain secured for any other employee except the intended receiver. If the manager has access to the private keys of his employees, then he will be able to benefit of the above features. A cryptosystem that allows a third party (called the *escrow agent* and represented in the proposed scenario by the manager) to access the messages that had been encrypted under the users public keys is said to provide key escrow abilities.

Multiple cryptosystems with key escrow capabilities, some of them fully secure in the standard model, have been defined in the literature [1–9], [12]. We propose a key escrow asymmetric system which is based on the construction of ASGKA (Asymmetric Group Key Agreement), a notion recently introduced [13].

We extend our system to hierarchical structures with multiple levels. For the use case that we have mentioned above, this extension is necessary because very few organizational structures have a unique management level. The rest of the paper is organized as follows: Section 2 exemplifies some applicability scenarios, analyze the drawback of the trivial solution and presents our contribution in accordance with the existing work in the field. Section 3 introduces the proposed system and analyze its security and performance. The extended system for hierarchical structures is given in Section 4. Finally, in Section 5 we conclude.

2 Asymmetric Cryptosystems with Key Escrow Capabilities

A key escrow cryptosystem is a system in which the cryptographic keys are entrusted to a third party (i.e. kept in escrow). The trusted third party is called the *Escrow Agency (EA)* (or more particular, it is represented by an *escrow agent*). In case that a KGC (Key Generation Center) exists in the system, then the EA and KGC are the same entity.

It is not mandatory that the third party owns the exact private keys of the users, but it is mandatory to benefit of the same features (it must be able to decrypt the messages encrypted under the users public keys). This is the case that our scheme implements.

2.1 Applicability

We exemplify the applicability of a key escrow system in general:

1. Communication monitoring

Lately there is a strong demand for monitoring communication to combat crime. A key escrow system fully provides such an ability: the EA/KGC is able to decrypt any ciphertext encrypted under a private key of any user.

This could also be the example of a manager that is able to supervise the communication of his employees. The ability may be implemented in a hierarchical structure: a manager can supervise the activity of his team members, but in the same time he is supervised by a superior manager that holds the same abilities upon him and his team. For example, a team leader can only supervise the activity of his team, a department director can supervise the activity of all the teams and team-leaders in his department and the CEO can supervise the activity of the entire company.

2. Delegation of duties

A use case scenario that exemplifies the delegation of duties consists of the ability of the manager (which plays the role of the EA) to read (and further response to) his employees e-mails while they are in holiday.

3. Backup of keys

In some schemes, it may be possible for the EA/KGC to reconstruct the secret key of the users. So, in case a user looses his private key, the EA may compute it and resend him the key. Therefore, the key escrow systems may provide backup in case of disaster.

Of course, key escrow ability is only applicable for systems that implement it. It is extremely hard to restrict the access of group members to any other encryption system that does implement this capabilities, making the communication unavailable to the EA. However, this can be achieved by considering all the other ways of communication illegal. In case of a company for example, the security policy of the organization can limit the encryption system to the desired one and penalize an employee in case of deviation.

The concept of communication supervision can also be seen from a malicious perspective. Consider the case of a person using a secure asymmetric communication system and a public-secret key pair generated by an apparently trusted authority. The malicious authority could have computed the corresponding keys so that it can decrypt all the ciphertexts using his own secret key, but without the users knowledge that they are under surveillance. This could be considered an example of SETUP (Secretly Embedded Trapdoor with Universal Protection), because it gives the apparently trusted authority the huge advantage of decrypting all the communication [11].

2.2 Trivial Approach

If the escrow agent owns all the private keys of the users, then he is able to decrypt any message that they receive. In case of systems that use a KGC, the scenario is trivial: the KGC generates a public-private key pair for each user, securely sends the private key to the corresponding user, but maintains a copy.

However, this approach has a major drawback. The quantity of information that the EA/KGC must hold depends on the cardinal of the group. Let's consider the case of a CEO of an important company. It is very difficult to keep the keys of all the employees secured in the same place. Besides the significant capacity of storage that is needed, the difficulty of keeping a secret raises with the quantity of information.

In our cryptosystem, the manager stores a quantity of information that is proportional with the depth of the hierarchical structure. Taking into consideration that an organizational structure is more wider than deep and the number of hierarchical levels usually does not exceed 5, the system becomes practical.

2.3 Related Work and Our Contribution

A special class of public key cryptosystems that provide key escrow capabilities are the IBE (Identity Based Encryption) schemes. The notion was first introduced by Shamir in 1984 as a type of public key encryption that allows a party to encrypt a message using the recipient's identity as public key [10]. Boneh and Franklin defined in 2001 the first efficient scheme and proved its security in the Random Oracle Model [2]. Improved results were obtained in the following years. In 2003 Canetti et al. proposed a scheme that was proved secure in a weaker security model (the Selective-ID Model) [6]. The result was further improved by Boneh and Boyen [3]. In the same year, Boneh and Boyen described a scheme that was proved to be fully secured in the standard model [4]. However, it was inefficient in practice. Their construction was slightly modified by Waters in order to become efficient while maintaining the same level of security [12].

Gentry and Silverberg introduced the first Hierarchical IBE [7]. Their construction is secure in the Random Oracle Model. In 2005, Boneh at al. described a scheme secure in the Selective-ID model [5]. One year later, Au et al. presents a HIBE fully secure in the standard model [1].

Our construction is not an IBE scheme and therefore it does not benefit of the IBE's specific advantages. However, our construction and the IBE schemes are similar in the sense that they are both asymmetric encryption schemes that provide key escrow capabilities. For more information about IBE, the reader may refer to the previous mentioned papers.

The purpose of our work is far from finding a better solution than the existing IBE solutions for the key escrow problem, but to build a solution based on the concept of asymmetric multi-key cryptography. In such a cryptosystem, more private keys correspond to each public key. Therefore, if a message is encrypted under a public key, then it can be decrypted by using any of the corresponding private keys.

The usage of asymmetric multi-key constructions for a key escrow system gives the EA the possibility to own a different secret key than the user, while maintaining the same capabilities of decryption. In case that one of the private keys is the same for all public keys of the users, then the escrow agent will store a single common key, instead of one key for each user. The storage capacity is significantly reduced in case of large groups. More, even if the escrow agent does not know the precise public key under which the encryption is performed, he is still able to decrypt efficiently.

Cryptosystems that fulfill this request were previously introduced in the literature. As an example, we mention the work of Boneh and Franklin that enables global escrow capability for the ElGamal encryption system [2] and Waters IBE [12].

Recently, Wu et al. introduced the notion of ASGKA (Asymmetric Group Key Agreement) as a primitive that permits more entities to establish a common group encryption key, but allow each member to hold different decryption keys [13]. A message encrypted under the group public key can only be decrypted by the group members, as they are the only owners of a corresponding private key.

It is easy to see that when the group is formed by exactly two users the construction of Wu et al. leads to two different private keys that correspond to the same public key. This is the exact situation we have mentioned before (a private key for the user and a different private key for the EA) and therefore, their construction can be used as a solution to the key escrow problem. In order to benefit of the advantages provided by the existence of a single key for the escrow agent, we have conditioned the construction so that the second private key that correspond to each public key is common for all users. This constitutes a valid construction for the previously mentioned scenario, but implies the mandatory existence of a KGC (the users do not generate their own keys, as in the ASGKA construction, but securely receive them from a trusted KGC).

3 The Proposed System

3.1 System Description

As we have explained in the previous section, the system uses two different private keys for each public key. One private key is common to all the public keys and belongs to the escrow agent, denoted as *the group manager* for the rest of the paper. The other private key belongs to the user.

The proposed system is based on the idea of the asymmetric group key agreement [13, 14] and uses the notion of bilinearity:

Definition 1. Be G and G_T two (multiplicative) cyclic groups of prime order p and g a generator of G. A bilinear map is a map $e : G \times G \to G_T$ with the following properties:

- 1. Bilinearity: for all $u, v \in G$ and $a, b \in \mathbb{Z}$, we have: $e(u^a, v^b) = e(u, v)^{ab}$.
- 2. Non-degeneracy: $e(g,g) \neq 1$.

We say that G is a bilinear group if the group action in G can be computed efficiently and there exists both a group G_T and an efficiently computable bilinear map $e: G \times G \to G_T$ as above [3].

Let $\{P_1, P_2, ..., P_n\}$ be the group of n participants (the users) and P_{man} the group manager (or more general, the key escrow agent). *PairGen* is an algorithm that receives as input a security parameter 1^{λ} and outputs a tuple $Y = (p, G, G_T, e)$, where G and G_T have the prime order p, g is a generator of G and $e : G \times G \to G_T$ is an efficient and non-degenerate bilinear map. *GenAuth* is the authority in charge with the keys generation (the KGC).

Construction 1. The Basic System

- 1: Phase 1. Public parameters generation
- 2: $PairGen(1^{\lambda}) \to Y = (p, G, G_T, e)$
- 3: Phase 2. Manager's keys generation
- 4: GenAuth randomly generates $h_{man}, X_{man} \in G, r_{man} \in \mathbb{Z}_p^*$ and computes:
- 5: $R_{man} = g^{-r_{man}}$
- 6: $A_{man} = e(X_{man}, g)$
- 7: $\sigma_{man} = X_{man} h_{man}^{r_{man}}$
- 8: Phase 3. Users keys generation
- 9: for $i = 1 \rightarrow n$ do
- 10: GenAuth randomly generates $h_i \in G$, $r_i \in \mathbb{Z}_p^*$ and computes:
- 11: $X_i = X_{man} h_{man}^{r_{man} r_i}$
- 12: $R_i = g^{-r_i}$
- 13: $A_i = e(X_i, g)$
- 14: $\sigma_i = X_i h_i^{r_i}$
- 15: end for

16: Phase 4. Users keys distribution

17: $\overline{GenAuth}$ publishes the public key of P_{man} : (R_{man}, A_{man}) and the public keys of the group members: $(R_i, A_i), i = 1..n$

18: GenAuth securely transmits the corresponding secret keys to P_{man} : (h_{man}, σ_{man}) and to each group member P_i : $(h_i, \sigma_i), i = 1..n$

```
19: Encryption
```

20: A message $m \in G_T$ is encrypted using the public key (R_i, A_i) of a group member P_i $(1 \le i \le n)$ as:

$$c = (c_1, c_2, c_3) = (g^t, R_i^t, mA_i^t)$$

where $t \in \mathbb{Z}_p^*$ is randomly chosen at encryption

21: Decryption

- 22: The ciphertext $c = (c_1, c_2, c_3)$ is considered for decryption
- 23: if decryption is performed by P_i then

24:
$$m = c_3/(e(\sigma_i, c_1)e(h_i, c_2))$$

25: **else**

```
26: if decryption is performed by P_{man} then
```

```
27: m = c_3/(e(\sigma_{man}, c_1)e(h_{man}, c_2))
```

- 28: else
- 29: decryption cannot be performed
- 30: end if

```
31: end if
```

The computation of X_i , i = 1..n, in step 11 is necessary for the correctness of the manager decryption:

$$\sigma_{man} = X_{man} h_{man}^{r_{man}} = X_i h_{man}^{r_i}, \forall i = 1..n.$$
(1)

Theorem 1. The basic system is correctly defined, in the sense that only the intended group member and the manager are allowed to decrypt a ciphertext, while it remains secured for the others.

Proof. Let $c = (c_1, c_2, c_3) = (g^t, R_i^t, mA_i^t)$ be the encryption of the plaintext m using P_i 's public key (R_i, A_i) . Then:

1. P_i is allowed to decrypt the ciphertext c:

$$\frac{c_3}{e(\sigma_i, c_1)e(h_i, c_2)} = \frac{mA_i^t}{e(\sigma_i, g^t)e(h_i, R_i^t)} = \frac{me(X_i, g)^t}{e(X_i h_i^{r_i}, g^t)e(h_i, g^{-r_i t})} = \frac{me(X_i, g)^t}{(e(X_i h_i^{r_i}, g)e(h_i^{-r_i}, g))^t} = \frac{me(X_i, g)^t}{e(X_i, g)^t} = m$$
(2)

2. P_{man} is allowed to decrypt the ciphertext c:

=

$$\frac{c_3}{e(\sigma_{man}, c_1)e(h_{man}, c_2)} = \frac{mA_i^t}{e(\sigma_{man}, g^t)e(h_{man}, R_i^t)} = \frac{me(X_i, g)^t}{e(X_i h_{man}^{r_i}, g^t)e(h_{man}, g^{-r_i t})}$$

$$= \frac{me(X_i,g)^t}{(e(X_i h_{man}^{r_i},g)e(h_{man}^{-r_i},g))^t} = \frac{me(X_i,g)^t}{e(X_i,g)^t} = m$$
(3)

3. $P_j, j \neq i$ is not allowed to decrypt the ciphertext c because

$$e(\sigma_j, g^t)e(h_j, R_i^t) \neq A_i^t \tag{4}$$

Observe that the manager also owns a public key (generated similar to the public keys of the users). At this point, the messages encrypted under his public key can be decrypted only by himself. In the extension scheme, any of his managers in the hierarchy will be able to decrypt them.

The key generation is performed by a trusted entity *GenAuth* (the KGC), fact that impose that the group members are confident in its behavior. More, the existence of secured channels is mandatory for the key distribution phase. Thinking at the scenario of the employees under the guidance of a manager, these assumptions are very possible to be achieved.

As a remark, if the manager P_{man} (the key escrow) plays the role of the *GenAuth* (KGC), then the keys distribution to P_{man} (first half of step 18) is no longer needed.

Example 1. Consider the group $\{P_1, P_2\}$ led by the manager P_{man} . From the basic system, we have:

- (R_{man}, A_{man}) the public key and (h_{man}, σ_{man}) the private key of P_{man} ;
- (R_1, A_1) the public key and (h_1, σ_1) the private key of P_1 ;
- $-(R_2, A_2)$ the public key and (h_2, σ_2) the private key of P_2 ;

$$\begin{aligned} R_{man} &= g^{-r_{man}}; \quad R_1 = g^{-r_1}; \quad R_2 = g^{-r_2}; \\ A_{man} &= e(X_{man},g); \; A_1 = e(X_1,g); \; A_2 = e(X_2,g); \\ \sigma_{man} &= X_{man} h_{man}^{r_{man}}; \; \sigma_1 = X_1 h_1^{r_1}; \quad \sigma_2 = X_2 h_2^{r_2}; \end{aligned}$$

and $\sigma_{man} = X_1 h_{man}^{r_1} = X_2 h_{man}^{r_2} = X_{man} h_{man}^{r_{man}}$ holds.

3.2 System Analysis

Security Analysis. Under the assumption that a perfectly secure channel exists between the *GenAuth* and the group members, the private keys are transmitted secure. Furthermore, in the distribution phase, the *GenAuth* should sign the messages in order to authenticate their origin and mark them with a timestamp mechanism in order to avoid replay attacks. All this safeguards considerable decrease the probability of an active attack.

Hence, for the rest of this section, we will only consider passive attacks. We show the indistinguishability of the basic system against chosen plaintext attacks (Ind-CPA) under the Decisional Bilinear Diffie-Hellman(DBDH) assumption.

Definition 2. Decisional Bilinear Diffie-Hellman Problem (DBDH): Given g, g^a , g^b , g^c , distinguish $e(g,g)^{abc}$ from $Z = e(g,g)^z$, $z \in \mathbb{Z}_q$.

The DBDH assumption holds if no adversary can solve the DBDH problem with a non-negligible advantage.

Theorem 2. The basic system is Ind-CPA secure under the Decisional Bilinear Diffie-Hellman (DBDH) assumption.

Proof. We consider the following Ind-CPA game between the attacker \mathcal{A} and the challenger \mathcal{C} :

1. C runs the key generation algorithm with the scope of solving the DBDH problem for g, g^a, g^b, g^c and Z. Phase 1 of the proposed system remains unchanged: $PairGen(1^{\lambda}) \rightarrow Y = (p, G, G_T, e)$. Next, in Manager's key generation phase, C chooses $h_{man} = g^a$ from DBDH and r'_{man} and X_{man} randomly. The public key of the manager is computed as described in the system, with the modification that r'_{man} was randomly selected instead of r_{man} :

$$R_{man} = g^{-r_{man}} = g^{-(r'_{man}+b)} = g^{-r'_{man}} (g^b)^{-1}$$
(5)

$$A_{man} = e(X_{man}, g) \tag{6}$$

During phase 3 of the basic system, C randomly selects $h_i, r''_i \in G$, i = 1..n, and computes:

$$g^{r'_{i}} = \frac{g^{r'_{man} + r''_{i}}}{g^{b}} \tag{7}$$

The challenger does not know the value of r'_i , but he is able to compute the public key of P_i as:

$$R_i = g^{-r_i} = g^{-(r'_i+b)} = g^{-(r'_{man}+r''_i-b+b)} = g^{-(r'_{man}+r''_i)}$$
(8)

From step 11 of the system, X_i is given by:

$$X_{man}h_{man}^{r_{man}-r_{i}} = X_{man}h_{man}^{(r'_{man}+b)-(r'_{i}+b)} = X_{man}h_{man}^{r'_{man}-r'_{i}} =$$

$$= X_{man}h_{man}^{r'_{man}-(r'_{man}+r''_{i}-b)} = X_{man}h_{man}^{-r''_{i}+b} =$$

$$= X_{man}(g^{a})^{-r''_{i}+b} = X_{man}(g^{a})^{-r''_{i}}g^{ab}$$
(9)

Although X_1 cannot be computed either, A_i is:

$$A_i = e(X_i, g) = e(X_{man}(g^a)^{-r''_i}g^{ab}, g) = A_{man}e(g^a, g^{-r''_i})e(g^a, g^b)$$
(10)

The challenger publishes $(R_i, A_i), i = 1..n$.

- 2. The attacker \mathcal{A} selects a public key (R_i, A_i) and outputs his selection and a pair of plaintext messages m_0 and m_1 , associated to it.
- 3. C chooses a random bit $b \in \{0, 1\}$, computes a challenge

$$c = (c_1, c_2, c_3) = (g^c, R_i^c, m_b A_{man}^c e(g^a, g^c)^{-r_i''} Z)$$
(11)

and gives it to \mathcal{A} . The challenger is able to compute the values even though c is unknown:

$$R_i^c = \left(g^{-(r'_{man} + r''_i)}\right)^c = \left(g^c\right)^{-(r'_{man} + r''_i)} \tag{12}$$

$$A_{man}^{c} = e(X_{man}, g)^{c} = e(X_{man}, g^{c})$$
(13)

4. \mathcal{A} responds with a bit $b' \in \{0, 1\}$ and wins if b' = b. The challenge is well formed¹ if and only if $Z = e(g, g)^z = e(g, g)^{abc}$:

$$A_{man}^{c}e(g^{a},g^{c})^{-r_{i}^{\prime\prime}}e(g,g)^{abc} = \left(A_{man}e(g^{a},g^{-r_{i}^{\prime\prime}})e(g^{a},g^{b})\right)^{c} = A_{i}^{c} \qquad (14)$$

Hence, the challenger C has the same advantage to solve the DBDH problem as the attacker A has to win the game.

Some remarks that apply to the underlying construction [13] remain valid:

- 1. If the same value $t \in \mathbb{Z}_p$ is used multiple times for the same public key, a related message attack could become possible;
- 2. The scheme is message homomorphic is the sense that $Enc(m_1, t_1)Enc(m_2, t_2) = Enc(m_1m_2, t_1+t_2)$, where Enc(m, t) represents the encryption of the message m using the parameter t.

Dynamicity Analysis. The proposed system allows any member to join or leave the group at any time, without restrictions regarding the cardinal of the group (as long as G and G_T are large enough). Whenever a new member joins the group, he is assigned a public and a private key, generated as usual. Whenever a member leaves the group, the manager just announces its public key as invalid. Because each member has a different decryption key, the former's group member knowledge of his own past private key does not affect the system. The great advantage is that no other keys must be modified.

If the group manager changes it is mandatory to reset all the keys and resume the protocol from the beginning. Otherwise he will continue to decrypt all the ciphertexts.

Performance Analysis. Table 1 summaries the required number of elements in G and G_T needed for key storage or transmitted as a ciphertext. We also count the number of operations performed for encryption, decryption, keys generation and joining the group.

				G			G_T			
	G	G_T		Sub	Mul	Exp	Mul	Div	Exp	е
Public key	1	1	Encryption	0	0	2	1	0	1	0
Private key	2	0	Decryption	0	0	0	1	1	0	2
Ciphertext	2	1	KeyGen	n	2n+1	3n+2	0	0	0	n+1
			Join	1	2	3	0	0	0	1

Table 1. Required storage capacity and computational performance

¹ We have used the same approach as in [13].

4 System Extension to Hierarchical Structures

The proposed scenario becomes applicable in practice if it allows an extension to hierarchical structures with an arbitrary number of levels. We propose next such an extension and analyze its security and performance.

4.1 Extended System Description

We consider a hierarchical structure with m + 1 levels, where 0 the highest level in the hierarchy (the manager on level 0 is able to decrypt any message delivered to any participant in the hierarchy) and m the is the lowest (a participant on level m is only allowed to decrypt his own messages)². The proposed extended system will associate to each participant m public and m private keys so that:

- Each participant on level $l, 0 \le l \le m$ has the first m l secret keys equal to the first m l keys of the manager situated on level 0;
- Each 2 participants with the same first superior on level f have all the last f parameters and keys equal and equal to the ones of their first common superior.

We consider the following notations:

- 1. **BasicGen**: phases 1 4 (steps 1-18) of the basic system for the manager on level 0 and phases 3-4 (steps 8-18) for any other manager (excluding P_{man} key distribution);
- 2. **FwdGen**: the BasicGen with a modification in step 10: *GenAuth* randomly generates $r_i \in G$ and sets $h_i = h_{man}$. This leads to $\sigma_i = \sigma_{man}$ (step 14), i = 1..n;
- 3. CopyGen: BasicGen with a modification in step 10: GenAuth sets $h_i = h_{man}, r_i = r_{man}$. This leads to $X_i = X_{man}, R_i = R_{man}, A_i = A_{man}, \sigma_i = \sigma_{man}$ (steps 11-14), i = 1..n.

Construction 2. The Extended System

1: Keys generation and distribution

```
2: for each level l = 0 \rightarrow m - 1 do
```

```
3: for each participant on level l do
```

```
4: for each key i = 1 \rightarrow m do
5: if i \leq m - l - 1 then
```

```
6: FwdGen:
```

```
7: else if i = m - l then
```

```
8: BasicGen
```

```
9: else
```

```
10: CopyGen
```

```
11: end if
```

 $^{^2}$ Do not confuse the plaintext m with the number of levels, also denoted by m. The significance is clear from the context.

12: end for

- 13: end for
- 14: end for
- 15: Encryption
- 16: A message $m \in G_T$ is encrypted using the public key $((R_{i1}, A_{i1}), ..., (R_{im}, A_{im}))$ of a group member P_i $(1 \le i \le n)$ as:

 $c = (c_1, c_{21}, ..., c_{2m}, c_3) = (g^t, R_{i1}^t, ..., R_{im}^t, m(A_{i1}...A_{im})^t),$ where $t \in \mathbb{Z}_p^*$ is randomly chosen at encryption.

17: Decryption

- 18: The ciphertext $c = (c_1, c_{21}, ..., c_{2m}, c_3)$ is considered for decryption
- 19: if decryption is performed by P_i then
- 20:

$$m = \frac{c_3}{e(\sigma_{i1}, c_1)...(\sigma_{im}, c_1)e(h_{i1}, c_{21})...e(h_{im}, c_{2m})}$$

21: else

22: if decryption is performed by P_{man} , any manager of P_i in the hierarchy, then

23:

$$m = \frac{c_3}{e(\sigma_{man1}, c_1)...e(\sigma_{manm}, c_1)e(h_{man1}, c_{21})...e(h_{manm}, c_{2m})}$$

- 24: else
- 25: decryption cannot be performed
- 26: end if
- 27: end if

Some improvements are possible (for example, set $A_i = A_{i1}...A_{im}$), but we prefer this form due to a simpler form of expression.

Theorem 3. The extended system is correctly defined, in the sense that only the intended group member and any of his managers in the hierarchy are allowed to decrypt a ciphertext, while it remains secured for the others.

Proof. Let P_i (on level l_i) be the intended receiver, P_j (on any level l_j , $l_j < l_i$) be one of his managers in the hierarchy and P_y , $y \neq i$ (on level l_y) a participant which is not a manager of P_i . Let $((R_{x1}, A_{x1}), ..., (R_{xm}, A_{xm}))$ be the public key of P_x and $((h_{x1}, \sigma_{x1}), ..., (h_{xm}, \sigma_{xm}))$ be the private key of P_x , $x \in \{i, j, y\}$.

1. P_i is able to decrypt a ciphertext encrypted with his own public key, because the following equation holds from construction:

$$e(\sigma_{ik}, g)e(h_{ik}, R_{ik}) = A_{ik}, \forall k = 1..m$$

$$(15)$$

2. P_i is able to decrypt P_i 's messages only if for all k = 1..m:

$$e(\sigma_{jk}, g)e(h_{jk}, R_{ik}) = A_{ik} \tag{16}$$

If $k \leq m - l_i - 1$ then $h_{ik} = h_{jk}$ and $\sigma_{ik} = \sigma_{jk}$ from FwdGen, so the equation holds.

If $m - l_i - 1 < k \leq m - l_j$, then CopyGen is performed on all levels lower that m - k, BasicGen is performed on level m - k and FwdGen is performed on all levels higher than m - k:

$$e(\sigma_{jk}, g)e(h_{jk}, R_{ik}) =$$

$$fromFwdGen: e(\sigma_{m-k-1k}, g)e(h_{m-k-1k}, R_{ik}) =$$

$$fromBasicGen: e(\sigma_{m-kk}, g)e(h_{m-kk}, R_{ik}) =$$

$$fromCopyGen: e(\sigma_{ik}, g)e(h_{ik}, R_{ik}) = A_{ik}$$
(17)

If $k > m - l_j$ then all parameters and keys are equal from CopyGen, so the equation holds;

3. P_y is not allowed to decrypt the ciphertext c, because from the BasicGen of the manager on level 0, holds at least:

$$e(\sigma_{ym}, g)e(h_{ym}, R_{im}) \neq A_{im} \tag{18}$$

Example 2. Consider the hierarchical structure from Figure 2. Because m = 2, each participant holds 2 pairs as his public key and 2 pairs as his private key. For clearness, we use a different notation.

At level l = 0, FwdGen and then BasicGen are performed:

$$\begin{array}{l} r_{123}, t_{123} \\ X_{123}, Y_{123} \\ h_{123}, k_{123} \\ R_{123} = g^{-r_{123}}, T_{123} = g^{-t_{123}} \\ A_{123} = e(X_{123}, g), B_{123} = e(Y_{123}, g) \\ \sigma_{123} = X_{123}h_{123}^{r_{123}}, \gamma_{123} = Y_{123}k_{123}^{t_{123}} \\ \end{array}$$

- $-((R_{123}, A_{123}), (T_{123}, B_{123}))$ is the public key of P_{123} ;
- $((h_{123}, \sigma_{123}), (k_{123}, \gamma_{123}))$ is the private key of P_{123} ;
- $-((R_{12}, A_{12}), (T_{12}, B_{12}))$ is the public key of P_{12} ;
- $-((h_{123},\sigma_{123}),(k_{12},\gamma_{12}))$ is the private key of P_{12} ;
- $-((R_3, A_3), (T_3, B_3))$ is the public key of P_3 ;
- $-((h_{123},\sigma_{123}),(k_3,\gamma_3))$ is the private key of P_3 .

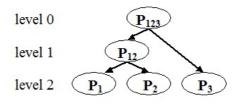


Fig. 1. The hierarchical structure used in Example 2

From BasicGen, we have: $\gamma_{123} = Y_{123}k_{123}^{t_{123}} = Y_{12}k_{123}^{t_{12}} = Y_3k_{123}^{t_3}$. At level l = 1, BasicGen and CopyGen are performed:

$$\begin{array}{ll} r_1,t_{12} & r_2,t_{12} \\ X_1,Y_{12} & X_2,Y_{12} \\ h_1,k_{12} & h_2,k_{12} \\ R_1 = g^{-r_1},T_{12} = g^{-t_{12}} & R_2 = g^{-r_2},T_{12} = g^{-t_{12}} \\ A_1 = e(X_1,g),B_{12} = e(Y_{12},g) & A_2 = e(X_2,g),B_{12} = e(Y_{12},g) \\ \sigma_1 = X_1h_1^{r_1},\gamma_{12} = Y_{12}k_{12}^{t_{12}} & \sigma_2 = X_2h_2^{r_2},\gamma_{12} = Y_{12}k_{12}^{t_{12}}, \text{ where:} \end{array}$$

- $-((R_1, A_1), (T_{12}, B_{12}))$ is the public key of P_1 ;
- $-((h_1, \sigma_1), (k_{12}, \gamma_{12}))$ is the private key of P_1 ;
- $-((R_2, A_2), (T_{12}, B_{12})) \text{ is the public key of } P_2; \\ -((h_2, \sigma_2), (k_{12}, \gamma_{12})) \text{ is the private key of } P_2;$

From construction, we have: $\sigma_{123} = X_{12}h_{123}^{r_{12}} = X_1h_{123}^{r_1} = X_2h_{123}^{r_2}$. Let $c = (c_1, c_{21}, c_{22}, c_3) = (g^t, R_1^t, T_{12}^t, mA_1^tB_{12}^t)$ be the encryption of a message m using P_1 's public key. Ciphertext c can be decrypted by:

– The receiver himself, P_1 :

$$\frac{c_3}{e(\sigma_1,c_1)e(\gamma_{12},c_1)e(h_1,c_{21})e(k_{12},c_{22})} = \frac{mA_1^tB_{12}^t}{e(X_1h_1^{r_1},g^t)e(Y_{12}k_{12}^{t_12},g^t)e(h_1,R_1^t)e(k_{12},T_{12}^t)} = \frac{mA_1^tB_{12}^t}{e(X_1h_1^{r_1},g^t)e(Y_{12}k_{12}^{t_{12}},g^t)e(h_1,g^{-r_1t})e(k_{12},g^{-t_{12}t})} = m$$

– The direct manager P_{12} :

$$\frac{\frac{c_3}{e(\sigma_{123},c_1)e(\gamma_{12},c_1)e(h_{123},c_{21})e(k_{12},c_{22})}}{\frac{mA_1^tB_{12}^t}{e(X_1h_{123}^{r_1},g^t)e(Y_{12}k_{12}^{t_{12}},g^t)e(h_{123},R_1^t)e(k_{12},T_{12}^t)}} = \frac{mA_1^tB_{12}^t}{e(X_1h_{123}^{r_1},g^t)e(Y_{12}k_{12}^{t_{12}},g^t)e(h_{123},g^{-r_1t})e(k_{12},g^{-t_{12}t})}} = m$$

– The manager P_{123} :

$$\frac{\overline{e(\sigma_{123},c_1)e(\gamma_{123},c_1)e(h_{123},c_{21})e(k_{123},c_{22})}}{mA_1^tB_{1_2}^t} = \frac{mA_1^tB_{1_2}^t}{e(X_1h_{123}^{r_1},g^t)e(Y_{12}k_{123}^{r_{12}},g^t)e(h_{123},R_1^t)e(k_{123},T_{1_2}^t)} = mA_1^tB_{1_2}^t}$$

 P_3 is not able to decrypt the ciphertext c because: $e(\gamma_3, g)e(k_3, T_{12}) \neq B_{12}$

Table 2. Required storage capacity and computational performance for the extended system (*if A_i is used instead of $A_{i1}...A_{im}$)

	G	G_T		G	G_T			
Public key	m	$m(1^*)$		Exp	Mul	Div	Exp	
Private key	2m	0	Encryption	m+1	$m(2^{*})$	0	1	
Ciphertext	m+1	1	Decryption	0	2m-1	1	0	2

4.2 Extended System Analysis

Security Analysis. We analyze the security of the extended system.

Theorem 4. The extended system is as secure as the basic system.

Proof. At each level, for each manager, BasicGen generates a key under the same conditions of the basic system. Hence one key from existing m (in a m + 1 level hierarchy) is as secure as in the basic system and the m keys are independent from each other, the proof is complete.

Dynamicity Analysis. If a new member without any subordinates joins or leaves the group, the situation is the same as for the basic system.

Whenever a manager leaves the group, the whole subtree under his position (including his own keys) must be resumed because otherwise he will continue to access his former subordinates encrypted messages. Since no modification is needed for the rest of the hierarchy, the proposed solution is of maximum efficiency regarding the number of keys that are altered.

Whenever a new manager joins the group, new keys must be generated for his corresponding subtree. If the height of the hierarchy remains constant, there is no need to modify any other keys. In case of increasing the number of levels, all existing keys must be completed to reach a number equal to the new height.

Performance Analysis. The required storage capacity and the computational performance of the proposed extended system for m + 1 levels are analyzed in Table 2.

5 Conclusions

We have proposed a system based on the construction of Wu et al. that implements key escrow capabilities. Considering the applicability of the scenario in the real world, we have extended our system to permit hierarchical structures. Both versions of the system were analyzed and proved to be Ind-CPA secure and efficient. Acknowledgements. This paper is supported by the Sectorial Operational Program Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number SOP HDR/107/1.5/S/82514.

References

- 1. Au M.H., Liu J.K., Yuen T.H., Wong D.S.: Practical Hierarchical Identity Based Encryption and Signature Schemes without Random Oracles. IACR Cryptology ePrint Archive 2006, 368 (2006)
- Boneh D., Franklin M.: Identity-Based Encryption from the Weil Pairing. Extended abstract in Crypto 2001, LNCS vol.2139, pp.213–229, Springer (2001)
- Boneh D., Boyen X., Goh E.: Efficient Selective-ID Secure Identity-Based Encryption without Random Oracles. Eurocrypt 2004, LNCS vol.3027, pp.223–238, Springer (2004)
- Boneh D., Boyen X.: Secure Identity Based Encryption without Random Oracles. Advances in Cryptology - CRYPTO 2004, LNCS vol.3152, pp.197–206, Springer (2004)
- Boneh D., Boyen X., Goh E. Hierarchical Identity Based Encryption with Constant Size Ciphertext. Eurocrypt 2005, LNCS vol.3494, pp.440–456, Springer (2005)
- Canetti R., Halevi S., Katz J.: A Forward-Secure Public-Key Encryption Scheme. Eurocrypt 2003, LNCS vol.2656, pp.482–499, Springer (2003)
- Gentry C., Silverberg A: Hierarchical ID-Based Cryptography. Asyacrypt 2002, LNCS vol.2501, pp.548–566, Springer (2002)
- 8. Horwitz J., Lynn B.: Toward Hierarchical Identity-Based Encryption. Eurocrypt 2002, LNCS vol.2332, pp.466–481, Springer (2002)
- Ren Y., Gu D.: Efficient Hierarchical Identity Based Encryption Scheme in the Standard Model. Informatica vol.32(2), pp.207–211 (2008)
- Shamir A.: Identity-Based Cryptosystem and Signature Schemes. Proceedings of CRYPTO 84 on Advances in cryptology, LNCS vol.196, pp.47–53, Springer (1985)
- Young A., Young M. The Dark Side of "Black-Box" Cryptography or: Should We Trust Capstone? Eurocrypt 1996, Proceedings of the 16th Annual International Cryptology Conference on Advances in Cryptology, LNCS vol.1109, pp. 89–103, Springer (1996)
- Waters B.: Efficient Identity-Based Encryption without Random Oracles. Eurocrypt 2005, LNCS vol.3494, pp.114–127, Springer (2005)
- Wu Q., Mu Y., Susilo W. Qin B., Domingo-Ferrer J.: Asymmetric Group Key Agreement. Proceedings of the 28th Annual International Conference on the Theory and Applications of Cryptography Techniques, LNCS vol.5479, pp.153–170, Springer (2009)
- Wu Q., Zhang X., Tang M., Peng Y., Zhenlong Q.: Extended Asymmetric Group Key Agreement for Dynamic Groups and Its Applications. China Communications vol.8(4), pp.32–40 (2011)
- Zhang L., Wu Q., Qin B., Domingo-Ferrer J., Gonzales-Nicolas U.: Asymmetric Group Key Agreement Protocol for Open Networks and its Applications to Broadcast Encryption. Computer Networks 55, pp.3246–3255 (2011)

A secure authenticated group key transfer protocol

Adrian Atanasiu¹ and Adela Georgescu¹

Faculty of Mathematics and Computer Science, Bucharest University, Str. Academiei 14, Bucharest 010014, Romania aadrian@gmail.com, adela@fmi.unibuc.ro

Abstract. In this paper we provide a different solution for the problem we introduced in a previous paper ([1]): a manager wants to build a team from a database of experts and establish a common secret key. We presented in that paper a proxy re-encryption-based solution expensive in terms of communication overhead (messages sent in the protocol). In the current paper, we present a more efficient solution, imposing also a new requirement on the protocol, the anonymity of the members of the team against each other. In contrast to the previous protocol, we propose a Diffie-Hellman key transfer protocol together with a security proof in a variant of the Bellare-Rogaway security model. Compared to other DH key distribution protocol, ours is better in terms of complexity of computation, but worse in terms of number of sent messages.

Key words: key transfer protocol, Diffie-Hellman, forward secrecy

1 Introduction

In the previous paper [1] we introduced the following problem: a manager wants to build a team of experts from a large database and establish a shared communication secret key with them. In this previous paper we presented a possible solution which has as underlying concept the *Identity-Based Proxy Re-Encryption* scheme. There were two initial conditions for that protocol:

- No one outside knows the list of experts chosen by the manager.
- The team members must agree on a commonly established key known only to them.

In this paper we developed a new solution for that problem to which we added a supplementary requirement. We impose the condition that only the manager knows the composition of the team, so we require anonymity of the selected members against each other, while in the previously proposed protocol [1] only anonymity against the audience was required. Our current protocol is a group key transfer protocol. Here, the manager computes a session key from some nonces he shares with the participants and transfers it securely and confidentially the secret key. The protocol we propose is different from the group key transfer protocols from the literature. The manager is a member of the his team, he computes the session key and he communicates with all the other members of the team in an mutually authenticated way, unlike the usual situation where each member of the group authenticates himself to a key generation center(KGC) in order to get the session key (see, for example, [9]). This latter approach requires that a trusted server to be set up, which increases the costs. We may say that our protocol is a group key transfer protocol with challenge response because of the messages sent from the manager to each member of the team.

The protocol we propose incorporates key authentication, confidentiality of the session key, forward secrecy and security against known session key. Also our protocol has a reduced number of rounds: only three. It is not endowed with key confirmation since this would require aditional rounds, but it can be achieved anyway, at this cost.

The paper is organized as follows: Section 2 presents some related work in the field of group key transfer protocols; in Section 3 we introduce some security attributes of our protocol and the cryptographic scheme used. Section 4 presents the construction of our protocol, while Section 5 discusses some general considerations and security issues. Some final conclusions are provided in last section.

2 Related work

Key establishment is a protocol whereby a shared secret becomes available to two or more parties, for subsequent cryptographic use. Key establishment may be broadly subdivided into key transport and key agreement [14]. Key transfer protocols depend on a mutually trusted key generation center (KGC) to select session keys and then transport session keys to all communication users secretly. In key agreement protocols, all communication users participate actively in computing the session key. The most commonly used key agreement protocol is Diffie-Hellman key agreement protocol [8].

Most distributed group key management protocols are based on the Diffie-Hellman key agreement protocol, such as, Ingemarsson et al. [11], Burmester and Desmedt [6], and Steiner et al. [15]. In 1996, Steiner et al. came with a natural extension of Diffie-Hellman protocol [15]. Bohli [4] developed a framework for robust group key agreement in 2006. One year later, Katz and Yung [12] proposed the first constant-round and fully scalable group Diffie-Hellman protocol. The fundamental security goal for a group Diffie-Hellman scheme to achieve is Authenticated Key Exchange (with "implicit" authentication), identified as AKE. In AKE, each player is assured that no other player aside from the arbitrary pool of players can learn any information about the session key.

3 Security

A key transfer protocol must meet a number of security attributes in order to be considered secure. Below we list some attributes necessary for our protocol:

- 1. security against known session key The security of the key transferred in an actual run of the protocol should not be compromised by an adversary who knows some old session keys.
- 2. forward secrecy Compromising long-term private keys of one or more parties should not affect the secrecy of previously established session keys.
- 3. key freshness A session key is fresh to a party A if A is assured that the key is a new value, rather than a reused value from a previous session.
- 4. **key authentication** assures an entity that only the other intended entities can possibly compute the session key.

The Diffie-Hellman Problem. Even if our protocol is a group key transfer protocol, all the flows are sent between the manager M and each participant. So we can view it, without losing generality, as two-party protocol where we employ the well-known Diffie-Hellman problem.

In a cyclic prime-order (q) group $\langle g \rangle$ generated by a g, we refer to the the following assumption: **The Decisional Diffie-Hellman(DDH) assumption**. Under this assumption, distinguishing g^{ab} from a random value when given g^a and g^b is computationally hard (a and b are random values from the group \mathbb{Z}_q where all the computations are made).

4 A general key agreement protocol

We present here our key transfer protocol that allows the manager M to securely distribute a communication key to his team of experts from the database. By $\mathcal{B} = \{P_1, ..., P_n\}$ we the database of experts from which M selects his team.

We are considering an open communication network where parties attempt to establish session keys with each other using a message-based key transfer protocol. We assume the existence of an algorithm for long-term key generation \mathcal{G} . So every member P_i of the database holds a pair of public/private keys (e_i, d_i) and he is identified by a public id_i . Let p be a large prime number. We work in the setting of the finite field \mathbb{Z}_p . The protocol runs as follows: In the first step, the manager invites all the experts to take part in the team, then he collects their answers, selects his team from those who accepted the request and sends them the session key in an authenticated and confidential way.

The security proof we provide for our protocols takes place in the random oracle model. All the participants in the protocol are supplied with two random functions

$$\mathcal{H}_1, \mathcal{H}_2: \{0, 1\}^* \longrightarrow \{0, 1\}^k$$

In instantiations, these functions will be modelled by some hash functions.

Setup the team

1. For every expert $P_i \in \mathcal{B}$, M chooses two nonces: α_i and a_i (both of them of k bits) which he sends signed with the private key. This step represents the request for participation addressed to all the experts in the database.

$$M \longrightarrow P_i : \{ id_M, \{\alpha_i\}_{e_i}, g^{a_i} \}_{sig_M}$$

for every $1 \leq i \leq n$.

2. If P_i accepts, then he generates a nonce b_i and sends the answer to M (signed with his private key)

$$P_i \longrightarrow M : \{id_i, \alpha_i, g^{a_i}, g^{b_i}\}_{e_M}$$

From the experts who accepted the request, M selects a team $T_M = \{P_{i_1}, \cdots, P_{i_k}\}, M \in T_M$. Let $I = \{i_1, \cdots, i_k\}$. He calculates for every member of the team (using the information received in step 2) the value $(g^{b_i})^{a_i} = g^{b_i a_i}$.

Transfer session key to the parties

3. M calculates the team session key

$$K = \sum_{j \in I} \alpha_j \pmod{p} \tag{1}$$

generates a nonce n_i (of k bits) P_i for every member of the database which answered at step 2 but was not selected in team and makes public some values m_i as follows:

For *selected* members with index $i \in I$, M publishes

$$\{kp_i = (K - \alpha_i) + g^{a_i b_i}, \mathcal{H}_1(g^{a_i b_i})\}_{sig_M}$$

and

for *unselected* members with index $i \notin I$ who answered in step 2, he publishes

$$\{kp_i = n_i + g^{a_i b_i}, \mathcal{H}_2(g^{a_i b_i})\}_{sig_M}$$

- 4. Every P_i who answered in step 2 proceeds as follows:
 - (a) calculates the value $(g^{a_i})^{b_i} = g^{a_i b_i}$ and checks the validity of the signature
 - (b) calculates the value $kp_i + \alpha_i$ which is the session key K if P_i was selected in the team or a different value if he was not selected.
 - (c) he checks if the second component of the message m_i is equal to the value $H_1(g^{a_ib_i})$; if this is true, then he was choosen in the team and the value $kp_i + \alpha_i$ represents the session key K; otherwise, if the second component is equal to $H_2(g^{a_ib_i})$ he was not chosen and he didn't get the correct session key.

5 Security analysis

A secure key agreement protocol should withstand two types of attacks: *passive attacks* - where the adversary can only observe the entitites carrying out the protocol - and *active attacks* where the adversary takes action by intercepting messages and replaying them, injecting false messages etc.

The adversary we model is afforded enormous power. She controls all communication between entities: she can read messages produced by them, provide messages of her own to them, modify messages before reaching destination, delay or replay messages. Furthermore, she may at any time initiate sessions between any two entities, engage in multiple sessions with the same party at the same time, and ask a party to enter a session with itself. As usual, the benign adversary is defined to be one that simply passes messages to and from the participants.

5.1 Description of the model

We should now define what it means for a protocol to be secure. Informally, an AKC protocol is secure, according to [3], if the following properties are met:

- no adversary can learn anything about a session key held by an uncorrupted party i provided that i has computed that session key in the belief that it is shared with another (uncorrupted) entity j.
- an accepting party i is assured that it has been involved in a real-time communication with j.

In the following, we introduce a formal security model for our protocol and then consider and prove in detail the security of our protocol in this model. We use a variant of the *Bellare-Rogaway* model for key establishment protocols, the same used in [3].

5.2 Security Model

The participants in the protocol are represented as a set of IDs. A player $U_i \in ID$ can have many instances modelled by oracles who are involved in different concurrent executions of the protocol. The instance s of player U_i is denoted by Π_i^s , with $s \in \mathbb{N}$. At any time during the execution of the protocol, an oracle may accept. When an oracle accepts it holds a session key sk, a session ID sid - an identifier used to uniquely name the current session and a pid which names the participant with which the instance believes it has just exchanged a key.

In the [3] model, every participant involved in the protocol is modelled by an oracle and can be addressed by an adversary through some allowed queries. An oracle Π_U^i represents the instance *i* of the participant *U*. A user *U* may have several instances involved in concurrent executions of the protocol. In order to express the concept of partnering oracle, we use session identifiers, as introduced in [2].

An adversary E may have access to all the oracles representing the participants in the protocol and she is allowed to make three types of queries to an oracle Π_U^s :

- $Send(\mathcal{O}, m)$. This query models message m being sent to oracle \mathcal{O} and returns an answer according to the execution of the protocol.
- $Reveal(\Pi_U^s)$. This query models the adversary's ability to find secret shared session keys. The oracle returns to the adversary the secret shared key K_s if he accepted it previously.
- Corrupt(U). This query corrupts a user U by revealing his long term secret sk_U . This query may be addressed only to an oracle who has accepted and no further action may involve any instance of user U.
- $Test(\Pi_U^s)$. This query models the semantic security of session key through the following game denoted by **Game**(\mathcal{A}, \mathbf{P}) between adversary \mathcal{A} and any of the oracles Π_U^s in the protocol P. During this game, \mathcal{A} may ask any of the above queries several times but only once the $Test(\Pi_U^s)$ query to a fresh oracle. Then, a random bit is chosen and returns the session key if b = 1 or a random value if b = 1. At the end of the game, adversary \mathcal{A} outputs a bit b' and wins the game if b = b'.

An adversary of the protocol is only allowed to address queries to fresh oracles, in order to avoid some trivial attacks. An oracle Π_U^i is *fresh* if it has accepted, it is uncorrupted, it remains unopened and there is no opened oracle $\Pi_{U'}^j$ which has the same session identifier as Π_U^i .

The security of the protocol is defined by the following game played **Game**(\mathcal{A}, \mathcal{C}) between the adversary \mathcal{A} and a challenger \mathcal{C} . The challenger \mathcal{C} generates the long term public and private keys for each user U by running the key ageneration algorithm on the security paremeter n. \mathcal{A} will interact with the oracles representing instances of the users through the queries defined above. At some point during the execution, a Test query is performed, and eventually, after other possibly addressed queries, A outputs a bit b' and terminates. The adversary wins the game if he is able to distinguish the session key obtained from the **Test** query from a random key, so if he guesses b' = b. We define $Good - guess^{\mathcal{A}}(k)$ to be the event that adversary \mathcal{A} correctly guesses whether b = 0 or b = 1, and the advantage of the adversary in outputting b' = b as

$$Adv^{\mathcal{A}}(k) = 2Pr[Good - guess^{\mathcal{A}}(k)] - 1$$

5.3 Definition of security

Definition 1. A shared secret key protocol is secure if

- 1. At the end of the execution of the protocol which is run in the presence of a benign adversary who simply relays messages, both oracles share the same session identifier and hold the same session key which is distributed uniformly at random over the key space.
- 2. For any adversary \mathcal{A} the advantage $Adv^{\mathcal{A}}(k)$ is negligible.

We now discuss the security of our protocol.

Theorem 1. Our protocol is a secure secret shared key protocol provided that the Diffie-Hellman assumption holds and the signature scheme involved in the protocol is secure.

Proof. The first condition from the definition of security follows imediately from the description of the protocol. We now concentrate on the second condition.

We prove in the following that for any adversary \mathcal{A} against the security of our protocol, there exists an adversary \mathcal{B} against the Diffie-Hellman problem and an adversary \mathcal{C} against the signature scheme Sig such that:

$$Adv^{\mathcal{A}}(k) < Adv^{DDH}(k) + Adv^{Sig}(k)$$

Let us assume that \mathcal{A} is an adversary against the security of our protocol. We define E to be the event that at the end of the game $\mathbf{Game}(\mathcal{A}, \mathcal{C})$ played between the adversary and the challenger, the oracle $\mathcal{O} = \Pi^i_U$ of adversary \mathcal{A} has on its transcript an incoming message $\{x_{\mathcal{O}}, g^{y_{\mathcal{O}}}\}_{sig_{\mathcal{O}^*}}$ that was not output by any other honest oracle in the game.

We have that

$$\begin{aligned} Pr[Game(\mathcal{A}, \mathcal{C}) = 1] &= Pr[Good - guess^{\mathcal{A}}(k) = 1] \\ &= Pr[Good - guess^{\mathcal{A}}(k) = 1 \cap E] \\ &+ Pr[Good - guess^{\mathcal{A}}(k) = 1 \cap \neg E] \\ &= Pr[Good - guess^{\mathcal{A}}(k) = 1 |E] \cdot Pr[E] \\ &+ Pr[Good - guess^{\mathcal{A}}(k) = 1 |\neg E] \cdot Pr[\neg E] \\ &< Pr[Good - guess^{\mathcal{A}}(k) = 1 |E] \\ &+ Pr[Good - guess^{\mathcal{A}}(k) = 1 |E] \\ &+ Pr[Good - guess^{\mathcal{A}}(k) = 1 |\neg E] \end{aligned}$$

We now construct two adversaries, an adversary \mathcal{B} against the DDH problem and an adversary \mathcal{C} against the employed signature scheme *Sig* according to weather or not event E occurs.

Let's suppose first that the event E does not occur; in this case, we construct algorithm \mathcal{B} as follows: \mathcal{B} is given as input the security parameter k and an instance of the decisional Diffie-Hellman problem ($\mathbb{G}, q, g, g^a, g^b, g^c$). \mathcal{B} plays the role of the challenger in the game **Game**(\mathcal{A}, \mathcal{B}). For this, B generates n identities in U and runs the key generation algorithm of the signature scheme Sig and that of the encryption scheme in order to obtain public/private key pairs for all the users in U. Adversary \mathcal{A} is allowed to make any of the Send, Corrupt, Reveal and Test queries to which algorithm \mathcal{B} answers as follows.

- If A makes a Corrupt(U) query, then \mathcal{B} returns sk_U and \mathcal{A} is no longer able to make any query to an instance of U.
- When \mathcal{A} makes a Send (\mathcal{O}, m) query to an oracle $\mathcal{O} = \Pi_U^i$, then algorithm \mathcal{B} answers in the following way:
 - Algorithm \mathcal{B} generates a random value $r_{\mathcal{O}}$.

- If U is an initiator then if message m asks him to initiate a session, \mathcal{B} returns to \mathcal{A} the message $\{id_{\mathcal{O}}, \alpha, (g^a)^{r_{\mathcal{O}}}\}$.
- Otherwise, for each oracle belonging to $pid_{\mathcal{O}}$ checks to find O^* such that the output of O^* is message m. Now \mathcal{B} can retrieve the value $r_{\mathcal{O}^*}$ and computes $(g^a)^{r_{\mathcal{O}}r_{\mathcal{O}^*}}$. In this way, \mathcal{B} collects the responses from every oracle belonging to $pid_{\mathcal{O}}$ and chooses a team for whose members he computes such a value $(g^a)^{r_{\mathcal{O}}r_{\mathcal{O}^*}}$.

Then \mathcal{B} computes the shared key K from all the values α that user U shared with other oracles: he computes $K = \sum_{j \in I} \alpha_j \pmod{p}$ where α_j is shared with an oracle \mathcal{O}_j that belongs to $pid_{\mathcal{O}}$. Then, depending whether the responder oracle was chosen in team or not, \mathcal{B} answers with one of the two messages: $\{(K - \alpha_i) + (g^a)^{r_{\mathcal{O}}r_{\mathcal{O}^*}}, (g^a)^{r_{\mathcal{O}}r_{\mathcal{O}^*}r_{\mathcal{O}^*}}\}_{sig_U}$ if \mathcal{O}^* was chosen in the team and $\{n_i + (g^a)^{r_{\mathcal{O}}r_{\mathcal{O}^*}}, (g^a)^{r_{\mathcal{O}}r_{\mathcal{O}^*}}\}_{sig_U}$ with n_i a nonce fresh generated if \mathcal{O}^* was not chosen in the team.

• Otherwise U is a responder; if he is in the **Setup** phase, then for each oracle belonging to $pid_{\mathcal{O}}$, \mathcal{B} checks to find \mathcal{O}^* such that \mathcal{O}^* output the message m. Next \mathcal{B} computes $(g^b)^{r_{\mathcal{O}}}$, obtains α and g^a and replies with $\{id_U, sk_{enc}(\alpha), g^a, (g^b)^{r_{\mathcal{O}}}\}_{sig_U}$.

If U is not in the **Setup** phase, then for each oracle belonging to $pid_{\mathcal{O}}$, \mathcal{B} checks to find \mathcal{O}^* such that \mathcal{O}^* output the message m. If he found \mathcal{O}^* , he checks that the signature from message m is valid, he retrieves the generated value $r_{\mathcal{O}^*}$ and computes $(g^a)^{r_{\mathcal{O}^r\mathcal{O}^*}}, (g^a)^{r_{\mathcal{O}^r\mathcal{O}^*}r_{\mathcal{O}^*}}$ and $(g^a)^{r_{\mathcal{O}^r\mathcal{O}^{r_{\mathcal{O}^*}}}$. He then checks the second component of message m, and depending on this value, he computes the session key from the first component.

- If a Reveal(O) query is made, then if \mathcal{O} accepted previously, \mathcal{B} returns the session key k to \mathcal{A} .

6 Comparison with related work

The protocol we introduce in this paper has the following charecteristics regarding efficiency: there are at most 3n messages sent over the three flows (step 1,2 and 3) from one execution of the protocol while the number of exponentiantions that each participant has to compute is at most 3.

We compare our results with related previous work. First note that our protocol different from all the other key establishment protocol regarding the way the shared session key is computed. In our protocol, one member of the team, the manager computes the key and transfers it to the rest of the team. Our protocol is more efficient in terms of the number of exponentiations than the protocol from [15] where this number is quadratic in n, the number of participants, but less efficient in terms of rounds and messages where they manage to achieve only n messages sent during the protocol. We remark that the latter protocol is rather a key agreement protocol since all the users take part in the computation of the session key. Compare now our protocol with [9]. In cited paper, they propose a key transfer protocol where the session key is chosen and distributed by a KGC which needs to share a secret key with each user before starting of the protocol, requiring the existence of a secured channel of communication with each participant. This protocol is not a DH protocol, therefore it's not based on exponentiations but on computing interpolations polynomials and requires 2n messages sent during one execution (provided that there exists additional secured channel of communication).

7 Conclusions

We built a group key transfer protocol where a manager sends a session key to his selected team of experts in a securely authenticated way. The protocol has a reduced number of rounds and it does not make use of a Key Generation Center. The confidentiality of the group key transfer phase of this protocol is information theoretically secure. We provide key authentication and anonymity of the members against each other, a property that has important applications in real life.

We have omitted the situation when two or several managers want to create their own team from the same database of experts, in the same time. We leave this for future work together with the situation where each of the two managers M_1 and M_2 of two simultaneous teams, from the same database, is a member in the team of the other one.

The efficiency of our protocol is better than [15] in terms of complexity of the operations involved in computing the session key but worse in the number of rounds and sent messages, at the cost of achieving anonymity, property that lead us to first executing a 2-party Diffie-Hellman key exchange which increases the number of sent messages during the execution of the protocol.

Acknowledgement

This work was sponsored by the European Social Fund, under doctoral and postdoctoral grant POSDRU/88/1.5/S/56668.

References

- 1. A. Atanasiu, Adela Mihaita. A key agreement protocol based on Identity-Based Proxy Re-encryption, Proceedings of the International Conference on Security and Management, 2011.
- M. Bellare, D. Pointcheval, P. Rogaway. Authenticated Key Exchange Secure Against Dictionary Attacks, Advances in Cryptology - Eurocrypt 2000 Proceedings, Lecture Notes in Computer Science Vol. 1807, B. Preneel ed, Springer-Verlag, 2000.
- S. Blake-Wilson, D. Johnson, A. Menezes. Key Agreement Protocols and their Security Analysis, the Sixth IMA International Conference on Cryptography and Coding, Cirencester, England, 1997.

- 4. J.M. Bohli A Framework for Robust Group Key Agreement, Proc. Intl Conf. Computational 16Science and Applications (ICCSA 06), pp. 355-364, 2006.
- 5. E. Brssot, O. Chevassut, D. Pointcheval, J.J. Quisquarter. *Provably authenticated group Diffie-Hellman key exchange*, CCS '01 Proceedings of the 8th ACM conference on Computer and Communications Security.
- M. Burmester and Y.G. Desmedt, A Secure and Efficient Conference Key Distribution System, Proc. Eurocrypt 94 Workshop Advances in Cryptology, pp. 275-286, 1994.
- 7. Colyn Boyd, Juan Manuel Gonzalez Nieto. Round-Optimal Contributory Conference Key Agreement Protocol.
- W. Diffie, M.E. Hellman, New Directions in Cryptography, IEEE Trans. Information Theory, vol. IT-22, no. 6, pp. 644-654, Nov. 1976.
- L. Harn, C. Lin, Authenticated Group Key Transfer Protocol Based on Secret Sharing, IEEE Trans. Computers, vol. 59, no. 6, pp. 842-846, June. 2010.
- C. Hsu, B. Zeng, Q. Cheng, G. Cui A novel Group Key Transfer Protocol, Cryptology ePrint Archive: Report 2012/043.
- I. Ingemarsson, D.T. Tang, and C.K. Wong, A Conference Key Distribution System IEEE Trans. Information Theory, vol. IT-28, no. 5, pp. 714-720, Sept. 1982.
- J. Katz and M. Yung, Scalable Protocols for Authenticated Group Key Exchange, J. Cryptology, vol. 20, pp. 85-113, 2007.
- C. Kudla, K. Paterson Modular Security Proofs for Key Agreement Protocols, ASI-ACRYPT 2005 2009/351.
- A. J. Menezes, P. C. van Oorschot and S. A. Vanstone, *Handbook of applied cryptography*, CRC Press, October 1996, pp. 490.
- M. Steiner, G. Tsudik, and M. Waidner, *Diffie-Hellman Key Distribution Extended* to Group Communication, Proc. Third ACM Conf. Computer and Comm. Security (CCS 96), pp. 31-37, 1996.

E-Voting algorithm dilemma: Blind signatures or Homomorphic encryptions

Pance Ribarski and Ljupcho Antovski

Faculty of Computer Science and Engineering, Skopje, Macedonia {pance.ribarski,ljupcho.antovski}@finki.ukim.mk

Abstract. More than twenty years are passing since the beginning of modern electronic evoting algorithms. Since then, two distinct paths are cleared as approach for evoting algorithm: Blind signatures and Homomorphic encryptions. Both of them have certain properties that are accounted as preferable, and both of them have weaknesses at certain points. In this paper we are discussing the pros and cons of both algorithms. We are implementing two candidates for both kinds and test them on various parameters. As conclusion we give points at when to choose Blind signatures, and when to choose Homomorphic encryptions.

Keywords: e-voting, blind signatures, homomorphic encryptions

1 Introduction

Scientist have been doing their work in the field of electronic voting algorithms for more than twenty years. Since the first publication from Chaum in 1982 [1] there have been many researchers that contribute to this field. There are many properties that these researchers added in their publications. Now the set of wanted properties that add to good features of a voting process exceeds the features that classic paper voting has.

Electronic voting is meant to be a replacement for classic paper voting. It should pertain the anonymity of voter-vote connection and should enjoy good faith with the voters. Electronic voting did not have a good start in big voting scenarios across the world. There are reportings about US voting processes that had gone bad. However, these bad examples of electronic voting implementations were not successful simply because they did not use algorithms from science work, or implemented them in the wrong way. In this paper we will discuss algorithms and the properties they have in voting processes.

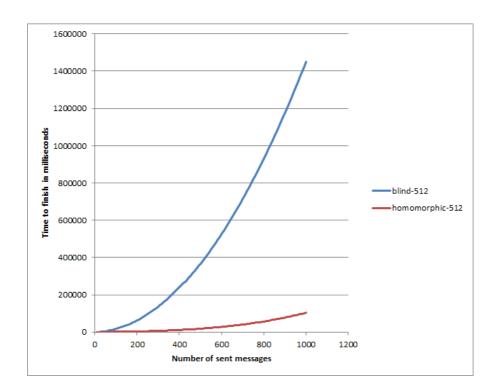
There are two major distinct paths in implementing electronic voting - Blind signatures and Homomorphic encryptions. Both of them have certain properties that are accounted as preferable, and both of them have weaknesses at certain points. In the following section we will describe the possible properties that electronic voting algorithms can fullfil. In the next two sections we will describe the two major types of electronic voting algorithms with explanation of representative algorithms. We have implemented these algorithms in Java in order to test them on various parameters. The concluding results will be given in the last section. This paper will give answers about choosing algorithms - which algorithm is better in which scenario.

2 Properties of evoting algorithms

There are many properties for electronic voting algorithms that scientist have contributed in their scientific work. Here we will list them in effort to standardize and unify the synonyms that appear in science literature.

- Correctness: This property is one of the base properties that one electronic voting system should have. When one algorithm satisfies this property it means that the tally at the end of the voting process is indeed the sum of all properly cast votes [2].
- Privacy: This property has roots in the classic paper voting. When everybody cast their vote in ballot boxes, nobody can trace the cast vote with the voter. This property is highly appreciated in electronic voting and also pretty hard to fullfil. When one algorithm satisfies this property it means that authorities nor other voters can trace the cast vote with the voter [3].
- Receipt-freeness: This property is recent in the science literature. It means that the voting system must not give proof of vote to the voter, or anybody else. This property is wanted when we want to prevent vote selling or coercing voters in any other way [3].
- Robustness: This property has its roots in software engineering and it is mostly wanted for large scale voting processes. It means that the electronic voting system should handle common errors and give the opportunity to continue the process when the conditions are fulfilled. The system must be prone to errors in the hardware, software, communication channels and any other participants in the system. This opportunity is generally connected with the term threshold robustness, which means that K-out-of-N entities are needed to correctly finish the process [4].
- Verifiability: When one algorithm has verifiability among its properties, it means that the voters or any other official bodies can verify the correctness and honesty of the system. Individual verifiability means that every voter can check if their cast vote is in the final tally. Universal verifiability means that anyone can check if the cast votes are counted properly [5].
- Eligibility: This property is also base property in voting processes. When one algorithm satisfies this property it means that only eligible voters can access the system and cast a vote. Usually this is done with voters' lists or pre-voting registration [6].
- Double votes: This property gives the chance that one voting system will detect double votes cast from authorized voters. The system should eliminate every duplicate vote. When fighting coercers the system may enable double votes and give a chance to the voter to later (or previosly) tell which is the correct vote [7].

- Fairness: This property is needed when the authorities do not want to give results during voting process. In some voting processes it is bad and can influence voters if they would know the current result. When one algorithm satisfies fairness it means that nobody knows the tallying result until the end of the process [5].



3 Blind signatures

Fig. 1. Blind signatures with Juang et al's algorithm compared to Homomorphic encryptions with Hirt and Sako's algorithm. On x-axis are the sent messages ranging from one to one thousand. On y-axis are the timings for each batch of votes given in milliseconds. The two algorithms are made with 512 bit keylengths.

The algorithms with Blind signatures are oldest in the electronic voting literature. They began in the 80ties when Chaum in 1982 [1] and in 1985 [8] for the first time used Blind signatures in electronic voting schemes. The general idea is that a voter sends encrypted vote to Administrator. Administrator can not reveal the vote, but can blindly sign it. Then the voter has blindly signed vote that can be sent to Counter. Counter checks if the vote is properly signed and decrypts and counts the vote. The basic idea can be explained with paper analogy as explained in [1]. Voter creates a ballot and inserts it into a special envelope with carbon. Then the sealed envelope is sent to Administrator. Administrator without opening the envelope signs it - thus creating a signature on the inserted ballot through the carbon envelope. The voter then opens the envelope and sends the blindly signed vote to Counter.

The cryptosystem used in Blind signatures is mainly RSA. Let's assume RSA with public modulus n and public exponent e. The private exponent is d. Voter chooses random r which is relative prime to n. Then he blinds his message m and gets his blind message m' as in (1).

$$m' = mr^e (mod \ n) \tag{1}$$

Because r is random it implies that m' is random too and doesn't leak any information about m. Administrator signs blindly this message by calculating (2).

$$s' = (m')^d (mod \ n) \tag{2}$$

The voter receives s' and reveals the blinding factor to reveal the signature s in (3).

$$s = s'r^{-1}(mod\ n) \tag{3}$$

There are many algorithms based on Blind signatures [1], [9], [10], [11], [12], [13], [5], [14]. In this paper we chose [14] as representative algorithm because it has many properties suitable for mass electronic voting processes. This algorithm solves the common problem of corrupt Administrator sending unused ballots.

3.1 Juang et al's algorithm

This algorithm is using scheme for Threshold blind signatures [15] which basically expands the Administrator role to more than one actor. This scheme uses secret sharing from [16]. With this secret sharing the group of Administrators is creating group public key, and when needed K-out-of-N Administrators can join to create the group private key.

The process of getting a signature involves K-out-of-N Administrators that must sign a vote. Then all these signatures combined together create the Blind signature of the group of Administrators. This neat property enables robustness where N-K Administrators can fail to work. Also this property takes the power from corrupt Administrators that can self-sign abstained votes and send them to Counter.

The algorithm in [14] is using another threshold group to achieve the fairness property. This group is made of N Scrutineers out of which K are needed by the [16] scheme for secret sharing. The voter encrypts his vote blindly signed by Administrators' group with the Scrutineers' public group key. At the end, at Publication phase K Scrutineers should join and give the private key for decrypting votes.

3.2 Results analysis

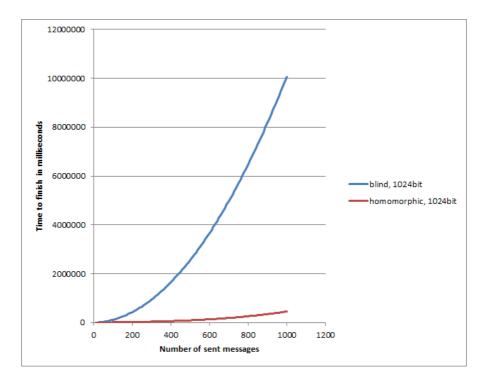


Fig. 2. Blind signatures with Juang et al's algorithm compared to Homomorphic encryptions with Hirt and Sako's algorithm. On x-axis are the sent messages ranging from one to one thousand. On y-axis are the timings for each batch of votes given in milliseconds. The two algorithms are made with 1024 bit keylengths.

The Blind signatures algorithm in [14] is extensive in using processor power. Because it involves many exponentiations in the calculations it is considered as very slow. The tests we made on messages (simulating votes) counting from one to one thousand on 512 bit keylengths are given in Figure 1. The tests with 1024 bit keylengths are given in Figure 2. In the tests with one thousand message and 512 bit keylength are 1451 seconds, and with 1024 bit keylength are 10057 seconds.

4 Homomorphic encryptions

The electronic voting algorithms based on Homomorphic encryptions pertain the same properties as Blind signatures. The Homomorphic encryptions satisfy the Fairness property by not decrypting votes until Count phase. The malleable property of some cryptosystem which leads to Homomorphic encryptions is classified as undesirable property. But cryptosystems which are malleable and enable Homomorphic encryptions are heavily used in electronic voting systems.

Homomorphic encryption for one cryptosystem means that algebraic operation over plaintext messages is equivalent with algebraic operations over ciphertext messages. Let M is the plaintext space and C is the ciphertext space. Let M is group under the operation \oplus while C is group over the operation \otimes . Then we say that ω is (\oplus, \otimes) homomorphic encryption if (4) holds.

$$E(m1) \oplus E(m2) = E(m1 \otimes m2) \tag{4}$$

The Homomorphic encryption is called partial if either addition or multiplication is allowed. If the cryptosystem allows both addition and multiplication then it is called fully homomorphic.

Examples for partial homomorphic encryptions:

- ElGamal: public key is (G, q, g, h) where $h = g^x$, x is private key and encryption of x is $\epsilon(m) = (g^r, mh^r)$ for random r from Z_q . Then the homomorphism is given in (5).

$$\epsilon(m_1)\epsilon(m_2) = (g^{r_1}, m_1h^{r_1})(g^{r_2}, m_2h^{r_2}) = (g^{r_1+r_2}, (m_1m_2)h^{r_1+r_2} = \epsilon(m_1m_2)$$
(5)

- RSA: public key is (m, e) and encryption of $x \in \epsilon(x) = x^e \mod m$. Then the homomorphism is given in (6).

$$\epsilon(x_1)\epsilon(x_2) = x_1^e x_2^e \mod m = (x_1 x_2)^e \mod m = \epsilon(x_1 x_2) \tag{6}$$

- Paillier: public key is (m, g) and encryption of x is $\epsilon(x) = g^x r^m \mod m^2$. Then the homomorphism is given in (7).

$$\epsilon(x_1)\epsilon(x_2) = (g^{x_1}r_1^m)(g^{x_2}r_2^m) = g^{x_1+x_2}(r_1r_2)^m = \epsilon(x_1+x_2 \mod m)$$
(7)

There are many algorithms based on Homomorphic encryptions: [17], [18], [19], [20], [21], [22], [3], [23], [24]. We chose the algorithm of [17] which is receipt-free modification of [24] or any other non-receipt-free algorithm.

4.1 Hirt and Sako's algorithm

Hirt and Sako in [17] present generic solution to many homomorphic algorithms which don't have the receipt-freeness property. We chose an algorithm which satisfies receipt-freeness because this property is important in voting processes where coercion and vote-selling is highly undesirable.

Receipt-freeness is very complicated to satisfy. If the system gives a choice to the voter to make an input, they can construct a message which is actual receipt or proof for voting. Hirt and Sako present a new way of creating choices - authorities create possible choices and permute them by sending them in some kind of a mixnet. The voter gets a list of choices, gets the permutations and can only point to the desired choice. That pointed choice is sent as vote to counting phase.

This algorithm is pretty fast thanks to the homomorphic property and relatively simple protocol between authorities. But the nature of creating votes as possible choices concludes to difficult process of getting the result. In 1-out-of-Lvoting where a voter chooses one candidate from a list with L candidates, every vote must be mapped to a specific number. Then from the sum of all cast votes, sums must be calculated for all candidates. In Hirt and Sako algorithm to get the result a dicrete logarithm calculation is needed which is supposed to be extremely difficult. There are known algorithms that are not brute force - such as baby step giant step [25]. Another way of finding the discrete logarithm is using precalculated hash tables. The problem size is reduced to partition of positive integers. This problem with supposed N voters and L candidates the possible combination and hash size is calculated as in (8).

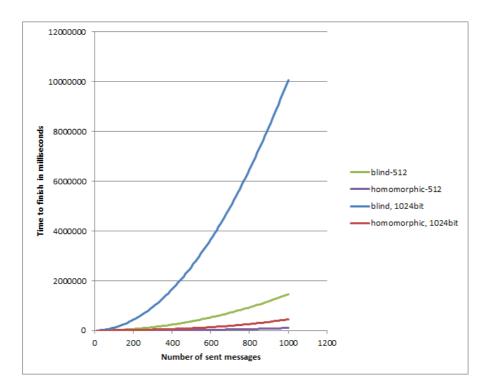


Fig. 3. Showing Blind signatures with Hirt et al's algorithm versus Homomorphic encryptions with Sako and Hirt's algorithm using 512 bit and 1024 bit keylengths. On x-axis are the sent messages ranging from one to one thousand. On y-axis are the timings for each batch of votes given in milliseconds.

$$\binom{N+L-1}{L-1} \tag{8}$$

4.2 Results analysis

Homomorphic encryptions are suitable for electronic voting systems. They are fast and pertain many good properties preferable for electronic voting. However, the problem with decoding votes is substantial. This problem arise in 1-out-of-L voting where L is rising and the number of voters N is also counting in the thousands. Results of Homomorphic encryptions algorithm with messages ranging from one to one thousand are given in Figure 1 and Figure 2. The two graphs show systems with 512 bit and 1024 bit keylength cryptosystems. The system with one thousand votes finished in 103 seconds with 512 bit and 452 seconds with 1024 bit keylengths.

5 Conclusion

The Figure 3 show unified timings for Blind signatures and Homomorphic encryptions with 512 and 1024 bit keylengths. Here we see that Homomorphic encryptions is substantially faster than Blind signatures. But, here the timings for Homomorphic encryptions don't include decryption and result calculations which we mentioned that is not easy task with many candidates and voters list count in the thousands. The conclusion would be that Homomorphic encryptions is perfect for voting processes with small candidate lists and small voters lists. If these numbers go high then Blind signatures is preferable choice. Both algorithms satisfy properties that are good for electronic voting - even mass electroning voting processes. Now practicioners and voting authorities should review and employ these algorithms in real voting scenarios. Voters should be introduced in the satisfied properties and should gain faith in these systems. The tests and experiments made in this paper show good results with given theoryonly algorithms. This is good proof for possible real-scenario systems. The evoting algorithm, be it Blind signatures or Homomorphic encryptions, takes a small part of any real evoting system. This paper contributes in the analysis towards choosing Blind signatures or Homomorphic encryptions. Comparing other parts of a real evoting system remains as a future work in our scientific work in the field of electronic voting systems.

References

- 1. Chaum, D.: Blind signatures for untraceable payments. In: CRYPTO. (1982) 199–203
- Benaloh, J.C., Yung, M.: Distributing the power of a government to enhance the privacy of voters. In: Proceedings of the fifth annual ACM symposium on Principles of distributed computing. PODC '86, New York, NY, USA, ACM (1986) 52–62

- Benaloh, J., Tuinstra, D.: Receipt-free secret-ballot elections (extended abstract). In: Proceedings of the twenty-sixth annual ACM symposium on Theory of computing. STOC '94, New York, NY, USA, ACM (1994) 544–553
- 4. Cohen, J.: Improving privacy in cryptographic elections. Technical report (1986)
- Fujioka, A., Okamoto, T., Ohta, K.: A practical secret voting scheme for large scale elections. In: Proceedings of the Workshop on the Theory and Application of Cryptographic Techniques: Advances in Cryptology. ASIACRYPT '92, London, UK, Springer-Verlag (1993) 244–251
- 6. Rjaskova, Z.: Electronic voting schemes (2002)
- Mu, Y., Varadharajan, V.: Anonymous secure e-voting over a network. In: Proceedings of the 14th Annual Computer Security Applications Conference, Washington, DC, USA, IEEE Computer Society (1998) 293–
- Chaum, D.: Security without identification: transaction systems to make big brother obsolete. Commun. ACM 28 (October 1985) 1030–1044
- Yaser Baseri, Amir S. Mortazavi, M.R.A.M.P., Mohajeri, J.: Double voter perceptible blind signature based electronic voting protocol. ISeCure, The ISC International Journal of Information Security 3(1) (2011) 1–8
- 10. Xia, Z., Schneider, S.: A new receipt-free e-voting scheme based on blind signature (abstract). (2006)
- Chou-CHen Yang, C.Y.L., Yang, H.W.: Improved anonymous secure e-voting over a network. Inf. Secur. v15 i2 (2004) 185–191
- Ray, I., Ray, I., Narasimhamurthi, N.: An anonymous electronic voting protocol for voting over the internet. In: In Proceedings of the Third International Workshop on Advanced Issues of E-Commerce and Web-based Information Systems. (2001) 21–22
- 13. Okamoto, T.: Receipt-free electronic voting schemes for large scale elections (1997)
- shenq Juang, W., laung Lei, C., twu Liaw, H.: A verifiable multi-authority secret election allowing abstention from voting. International Computer Symposium 45 (2002) 672–682
- Juang, W.S., Lei, C.L.: Partially blind threshold signatures based on discrete logarithm. Computer Communications 22(1) (1999) 73 – 86
- Pedersen, T.P.: Non-interactive and information-theoretic secure verifiable secret sharing. In: Proceedings of the 11th Annual International Cryptology Conference on Advances in Cryptology. CRYPTO '91, London, UK, Springer-Verlag (1992) 129–140
- Hirt, M., Sako, K.: Efficient receipt-free voting based on homomorphic encryption, Springer-Verlag (2000) 539–556
- 18. Huszti, A.: A secure electronic voting scheme, Periodica Polytechnica (2007)
- Acquisti, A.: Receipt-free homomorphic elections and write-in voter verified ballots, Carnegie Mellon University (2004)
- Damgard, I., Groth, J., Salomonsen, G.: The theory and implementation of an electronic voting system. In: SECURE ELECTRONIC VOTING, Kluwer Academic Publishers (2003) 77–100
- Baudron, O., Fouque, P.A., Pointcheval, D., Stern, J., Poupard, G.: Practical multi-candidate election system. In: Proceedings of the twentieth annual ACM symposium on Principles of distributed computing. PODC '01, New York, NY, USA, ACM (2001) 274–283
- Katz, J., Myers, S., Ostrovsky, R.: Cryptographic counters and applications to electronic voting. In: Proceedings of Eurocrypt 2001. (2001) 78–92
- 23. Hirt, M.: Receipt-free -out-of- voting based on elgamal encryption. In: Towards Trustworthy Elections. (2010) 64–82

- 24. Cramer, R., Gennaro, R., Schoenmakers, B.: A secure and optimally efficient multi-authority election scheme, Springer-Verlag (1997) 103–118
- Shanks, D.: Class number, a theory of factorization, and genera. In: 1969 Number Theory Institute (Proc. Sympos. Pure Math., Vol. XX, State Univ. New York, Stony Brook, N.Y., 1969), Providence, R.I. (1971) 415–440

DSP Tools for Signal Measurement Performance Analysis

K. A. DOTCHE

Department of Telecommunications Engineering

Kwame Nkrumah University of Science and Technology

KNUST-Kumasi / Ghana

Email: kdotche2004@gmail.com

Abstract. This paper introduces a novelty method of using a Linear Prediction Coefficient (LPC) filter, a digital signal processing (DSP) tool to get accurate signal measurement in noisy mobile environment. By measuring the power of mobile radio, it also measures the coverage of an area served by several base stations. For results' validation, the received power at the mobile user end of two Code Division Multiple Acccess-2000 (CDMA2000) cellular networks operating at different frequency in the same environment, Lome in Togo, is been considered for the coverage analysis. Our analysis has shown that within the problem areas the filter response does not match with the measured data. These mismatching areas may result from poor soft-handoff process or a dead zone. The method has shown significant help in problem areas identification. Consequently, such a filter can be embedded to current firmware for Radio Frequency coverage optimization, and for effective spectrum efficiency.

Keywords: Radio channel, received power, coverage analysis, linear coefficient prediction (LPC), drive test.

1 Introduction

Radio coverage estimation is very costly for a mobile cellular operator. It may be admitted that it is the most perceived criterion of performance by mobile users. In order to add value and save their investment, the network performance may constantly be monitored. The cellular network coverage performance is thus impaired by the network design, the continuous optimization process and the surrounding clutter at both the Mobile Station (MS) and the Base Station (BS). The clutter constitutes the radio space channel. It is defined as the air interface between the Mobile Station (MS) and the Base Station to the extent of the terrain's roughness, etc. In this environment, the signal is heavily influenced by reflections from buildings. The clutter effect on the radio signal received at MS inflects the signal to vary randomly. This variation in received signal is quantified by the effect of shadowing or

environment fading which is mainly depended on the mobile user location (effect of population density).

The fading envelope of a received signal strength may be viewed as having two main variant components [1], the fast fading signal observed over small distance (known as Small Scale Fading: SSF) and slow fading signal as over a long distance (known as Large Scale Fading: LSF or Slow Fading). The large-scale fading is related to diffraction and reflection effects that the multipath components undergo on their way between the transmitter and the receiver. The small-scale fading or fast fading is rapid changes observed in received signal. These variants were caused by the user mobility [2]. The fast fading magnitude was obtained on an average number of finite sample measurements but the slow fading has consisted to carry out an average of wide measurement over long distances [1]. On the other hand, another general phenomenon, which is a source of signal attenuation, is multipath propagation. Multipath propagation has been analyzed by several stochastic models [3], [4] available in literature. When all multipath signals reach the receiver with same levels); in this particular case, there is no direct rays that reach the receiver, the Rayleigh fading model is applicable. Another type of channel fading is the Rice fading model. In this type of channel fading, one multipath signal is stronger than the rest; a direct ray may likely reach the receiver point. Thus, it is appropriate for line-of sight channel. Both fading models have been generalized in the Nakagami fading model [5]. However, the area of radio channel estimation remains broad to investigate; this includes the modern cellular network coverage analysis.

Cellular network coverage analysis is a delicate task for the Radio Frequency (RF) planning engineer at deployment stage as well as the optimization phases. Several tools for RF coverage analysis in cellular networks are available on the market. These tools, such as Agilent analyzers, cell-test, master-site, Airbridge and so on; perform channel measurement adequately. But then, it may be admitted that these tools perform statistical analysis of the radio network coverage either per base station or per sector with regard to a predefined threshold. Recently, radio channel estimation and its components analysis have gained attention in using a digital signal processing (DSP) tool to get accurate channel estimation as in [6]. The work has proposed a dynamic adaptive estimator which comprises the spline and smoothing spline tools to discriminate path loss components on received signal measurement, such as path loss, shadowing and fast fading. This method debunks from the statistical methods used path loss components estimation as in [7] and [8]. In [9], a communication channel tracking apparatus has been proposed which uses a linear predictor a DSP tool for partial channel estimation with finite time sample.

The study in this paper introduces a novelty method for cellular coverage analysis by using a Linear Prediction Coefficient (LPC) filter [10], a DSP tool to investigate the mobile handset received power level in noisy environment. The received power measurements have been performed on field installation of two different Code Division Multiple Access-2000 (CDMA-2000) cellular networks operating in the same environment in Lome, Togo. Our contribution remains in light of exploring dynamic filter for radio channel estimation.

The rest of the paper is structured as follows: the section 2 illustrates the linear predictor as coverage estimator and the theory behind, the data collection using the drive testing is presented in section 3; the comparative results between the estimator response and the measured data are been object in section 4 and the section 5 is devoted to the conclusion and future works.

2 The linear predictor (lpc) as Coverage estimator

In this section the theory of the shadowing as frequency dependency and lpc as the coverage estimator in noisy mobile environment are been addressed.

2.1 The received power and shadowing frequency dependency

The received power p_{R_X} at MS as function of the frequency [4] is as given in (1):

$$p_{R_X} = p_{T_X} * g_{T_X} * g_{R_X} * \left(\frac{c}{4*\pi}\right)^n * \left(\frac{1}{d}\right)^n * \left(\frac{1}{f}\right)^n * x_{\sigma}.$$
(1)

Let l (the path loss) define the ratio of transmitted power p_{T_X} to the received power p_{R_X} in watt scale. Though, by convention the path loss L(dB) is always expressed in dB scale as the observed difference in the base station transmitted power $P_{T_X}[dBm]$ and the received power $P_{R_X}[dBm]$ at mobile handset in mobile communication environment [3], [4]. Thus, we can write:

$$\begin{cases} l = -\frac{p_{R_x}}{p_{T_x}} \Rightarrow \\ L(dB) = -10 * \log_{10}(\frac{k}{f^n} * x_{\sigma}) \end{cases}$$
 (2)

where

$$k = g_{T_X} * g_{R_X} \left(\frac{c}{4*\pi}\right)^n * \left(\frac{1}{d}\right)^n.$$
(3)

$$K(dB) = G_{T_x} + G_{R_x} + 10nlog_{10}\left(\frac{c}{4\pi}\right) - 10nlog_{10}d$$
(4)

$$L(dB) = K - 10nlog_{10}(f) + X_{\sigma}(dB).$$
(5)

 p_{T_X} is the transmitting antenna power and transmitting antenna gain g_{T_X}

 g_{R_X} is the user handset (MS) antenna gain,

n the path loss exponent which describes the environment,

d distance of separation between a BS and an MS

 \mathbf{x}_{σ} is the shadowing effect of the environment;

 \mathbf{K} is a measure of the local mean variation in the large scale sense and is distance dependency;

 $\lambda = c/f$ is the wavelength, f is the carrier frequency, and c is the speed of light in vacuum;

L is the path loss having three dependency parameters f, d and n

If we assume that shadowing can be viewed as a correlated error on the measurement in a time varying process and can be derived as a frequency dependency function as shown in (6):

$$X_{\sigma} = L(dB) - K(dB) + 10n \log_{10} f$$
. (6)

It comes as in (7)

$$\begin{cases} X_{\sigma} \propto C + 10n.\log_{10} f \\ C, cst \end{cases}$$
(7)

 X_{σ} is therefore frequency and environment dependency.

2.2 The linear predictive coefficient or linear predictive coding (LPC)

The LPC determines the coefficients of a forward linear predictor by minimizing the prediction error in the least squares sense (Figure)1. It has applications in filter design, speech coding, and system identification [11]. In system analysis, linear predictive coding is known as a mathematical optimization process. In Matlab software [12], the function lpc(x,p) finds the coefficients of a *pth*-order linear predictor (FIR filter) that predicts the current value of the real-valued time series x based on past samples. Where, the predicted signal $\hat{x}(n)$ is given as in (8):

$$\hat{x}(n) = \sum_{i=1}^{p} a_i x(n-i)$$
 (8)

where *p* is the order of the prediction filter polynomial, and $a_i = [1, a(2) \dots a(p + 1)]$, is the predictor coefficient.

with the predictor error e(n) given as in (9) and (10)

$$e(n) = x(n) - \hat{x}(n)$$
(9)
= $x(n) - \sum_{i=1}^{p} a_i x(n-i)$
 $e(n) = -\sum_{i=0}^{p} a_i x(n-i)$. (10)

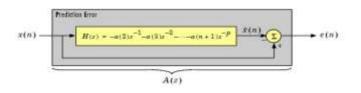


Figure 1 The LPC predictor block

It may be noted that a_i is the autocorrelation criterion. For $a_i = 1$ the predictor is linear filter.

The prediction error e(n) can be viewed as the output of the prediction error filter A(z) as shown in **Figure 1**, where H(z) is the optimal linear predictor, x(n) is the input signal.

2.3 Signal components and coverage estimation

In **Figure 2**, it is showing the linear adaptive filter incorporated with the LPC tools, to perform the radio signal measurement in cellular network coverage. For a given data set (mainly for a fixed transmitter), one may estimate the path loss components, and evaluate the coverage analysis as whole through the LPC filter.

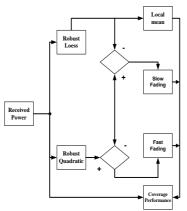


Figure 2 Linear adaptive filter

The estimating of the fast fading is achieved by computing the Robust Lowess (linear fit) algorithm. This is an advanced smoothing spline method. Meanwhile, the slow fading is obtained by using the Robust Loess (Quadratic fit) algorithm [13]. However, the local mean is obtained by subtracting the fast fading quantity out of the slow fading value.

3 Methodology of data collection

The measurement experiment has consisted of performing a simultaneous drive testing in two cellular networks operating at two different frequencies with two CDMA Handsets (Huawei C5320 for 450MHz and Kyocera Kx5 for 850MHz). The 450MHz network has known a recent extension. The second network of 850MHz is not loaded during the test and recently deployed. Both networks are operating in Lome and belong to the same operator (Togo Telecom). The setup of the measurement procedure is illustrated in **Figure 3**. The drive testing is carried out by a roaming in the wireless cellular network with a car equipped with over-air test (OAT) tools. These tools are: a Global Positioning System (GPS) receiver antenna placed in

the car rooftop (to obtain the accurate distance of location) and GPS 76 software version 2; a drive-test software (ZK CellTest tool [14] capable to to read the dongle (see Figure 3) and a communication network analyser software (here, the Gladiator firmware for extracting the call log traces after the drive-testing is completed, assigned in the office), a laptop (the client), Map info software (for loading the digital map); Inverter (for converting the DC source obtained from the car inner-port to the AC source with purpose of supplying power to the laptop in case of a long time drive testing); an Extension board (for connecting the laptop, and various hardware that need AC current).

The inverter in **Figure 3** is meant for converting the Direct Current (DC) voltage at the output of the connector to Alternative Current (AC) (220). This is used to supply sufficient power to the laptop during a whole day drive test. For very a few minutes' drive-testing, the inverter may not be necessary. The Mobile Equipment line (M1) will be calling the Mobile Equipment line (M2). In some cases, the line (M2) can be assigned to a fixed line office. In this case, one mobile may be used for the drive test.



Figure 3 drive test setup system

The drive test traces' map is showing in **Figure 4** as in multi-transmitters' system in a noisy mobile cellular environment.



Figure 4 Drive test road map

The measurement was a simultaneous and continuous in both operating frequencies of the transmitters, in 450MHz and 850MHz. The extracted data from the Gladiator firmware was processed to Matlab software and been analyzed by the usage of the lpc estimator.

4 Results and Discussion

For a fixed transmitter, one may use the robust linear fit method in 'cftool' in Matlab software for the estimation of the path loss components as shown in **Figure 5**.

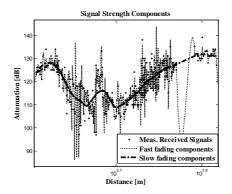


Figure 5 Received signal components in mobile cellular environment

In **Figure 5**, it is showing the path loss components obtained using the robust loess linear fit to quantify the slow fading on the measurement. We also used the robust quadratic to show the fast fading components on the measurement. The obtained plots show a better agreement with the measurement data as compared to the simple smooth spline method proposed in [6]. **Figures 6**, 7 and **8** are shown the received signal level being processed as a time series to the adaptive filter proposed in **Figure 2** for the coverage performance estimation.

In Figure 6, it shows the measured data compared to the filter response in both frequencies in a subplot comparison in 450MHz and 850MHz with their respective error response. In 850MHz the Linear PLC filter matches with the measurement as shown in Figure 6 and Figure 7. This may be explained by the fact that the transmitters in used are sufficiently enough for the intended coverage. Aside this, the load may say to be low. Therefore the system margin gain is seen to be reliable for the soft-handoff. The channel impairments such as the shadowing and multipath components are markedly compensated (reduced) because of the use of directional antennae and the down-tilt technique [15].

In 450MHz, the network may present some deficiency such as persistent dead zone thus dropped calls as illustrated in **Figure 6** and **Figure 8**. This may show an inconstancy of dimensioning at design process. Consequently it may be explained by the fact that on one hand, the distance of separation among Base Transceiver Subsystem often called Base Station (BTS/BS) is not well set thus the cell overlapping is large at some areas. Meanwhile on the other hand, this may also be viewed as due to the extension of the network. Particularly, adding new Pseudo Noise (PN) code [16] is also a delicate process to an existing network.

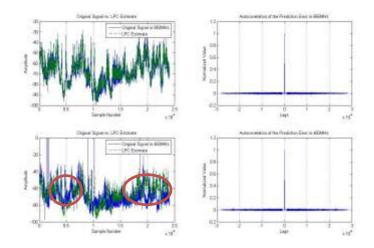


Figure 6 Measurement and LPC predicted data

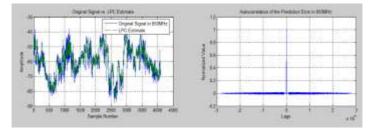


Figure 7 Sample view of snapshot in 850 MHz

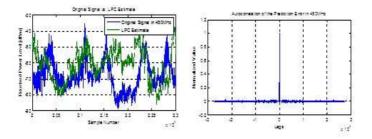


Figure 8 View of problem areas in 450MHz

A poor evaluation of PN may highly create interference in the network coverage area. These interference areas observed may result in unnecessary soft-handoff at times. This may also be explained by the size of soft-handoff window-search [17]. In such cases, if the size of the soft-handoff window-search is small then unnecessary soft-hand may also occur even at a relative short distance of the potential BTS. We may then argue that the small mismatching areas are zone of poor soft-handoff, whereas

the large mismatching areas may be resolved in an extremely poor hard handoff decision due to high interference from other non-CDMA antennae. Though the antennae' heights of 450MHz BTSs are mostly the highest as compared to others sharing the same mast but users at cell-edge may observe high interference. This situation may then lead to dropped call in this mismatching as compared to the filter response. These areas can be easily localized by the tracing log files.

Theoretically, it may also presume that the response of filter may show a little agreement in 850MHz which could reveal the effect of shadowing in high frequency. But then, the autocorrelation of the predictor error shows that, the shadowing may be viewed as an error on the received signal and may be approximated to a white Gaussian noise. Therefore a shadowing effect in cellular environment can also be modeled as a white Gaussian noise.

5 Conclusion

In this paper, we have shown the feasibility of RF network coverage performance analysis with help of a Digital signal processing tool, the LPC filter. The LPC filter often used in speech coding was used to analyze the received power from two CDMA2000 cellular networks operating at different frequency in the same environment. The analysis shows that within the problem areas the filter response does not match with the measured data. These mismatching areas may be the results of poor soft handoff dimensioning or a dead zone in the coverage area.

Such a tool can be embedded to current RF coverage analysis network tools, as a case for cognitive radio. This may give significant help in problem areas identification, such as dead zone and poor soft-handover in cellular network coverage. In the future, estimation of radio channel parameters such as the path loss and shadowing component will be carried out using LPC filter incorporated to a dynamic channel estimator proposed in [6].

Acknowledgments. We wish to express our gratitude, to the anonymous reviewers for their valuable contributions, to Dr K. Diawuo; to Dr Komlan Yovo, CTO of Togo-Telecom Ltd.; and Global-Tech France, for their technical support.

References

1. Lee, William, C.,Y.: Estimate of Local Average Power of a Mobile Radio. In: IEEE Transactions on Vehicular Technology, VOL. VT-34, No. 1, (1985)

2. Egli, John: Radio Propagation Above 40 MC Over Irregular Terrain. In: IEEE Transaction on Vehicular Technology, New York (1957)

3. Parsons, J., David: The Mobile Radio Propagation Channel. John Wiley and Sons, Second Edition, London (1992)

4. Rahnema, Moe: UMTS Network Planning, Optimization and Inter-Operation with GSM. John Wiley & Sons Asia Pte Ltd (2008)

5. Oestges et al.: Experimental Characterization and Modeling of Outdoor-to-Indoor and Indoor-to-Indoor Distributed Channels. Technical report, European Cooperation in the field of Scientific and Technical Research, Braunschweig (2009)

6. Le Cornec et al.: Optimization of Radio Measurements Exploitation in Wireless Mobile Networks. In: Journal of Communications - JCM, vol. 2, no. 7, pp.59-67(2007)

7. Adebayor et al.: Characterization of Propagation Path loss at 1.8 GHz: A Case Study of Benin-city, Nigeria. Journal of Applied Sciences, Medwell Online Vol 1(1-4) 92-96 (2006).

8. Konak, Abdullah: Estimating Path Loss in Wireless Local Area Network Using Ordinary Kriging. In: Proceedings of the 2010 Winter Simulation Conference,

Pennsylvania, USA (2010). http://www.informs-sim.org/wsc10papers/267.pdf.

9. Gifford et al.: A Communication Channel Tracking Apparatus. US7433434 B2 United States of America (2008).

10. Wikipedia Encyclopedia, http://en.wikipedia.org/wiki/Linear_prediction.

11. Jackson, L.B.: Digital Filters and Signal Processing. s.l. : Second Edition, Kluwer Academic Publishers, pp. 255-257 (1989)

12. MathWorks.: Matlab software version 8.

http://www.mathworks.com/help/techdoc/rn/f14-998197.html.

13. Ronaldo, Dias.: Nonparametric Regression: Lowess/Loess.

[http://www.ime.unicamp.br/~dias/loess.pdf]

14. ZK Celltest, http://www.zk.com/ (2009)

http://www.zk.com/support/release_info_new.html.

15. Dotche, Koffi, A. and Diawuo, Kwasi: Received Power in Antenna Downtilt in Cluttered Mobile Environments: A Case Study of West-Africa. ISBN:

9783639339697, VDM, London (2011)

16. Ganesh, Rajamani: Impact of adding sites on PN Offset Planning in CDMA

networks. GTE Laboratories Incorporated, USA, (2007)

17. Yang, C., Samuel: CDMA RF Engineering. Artech House, Inc, (1998)

Society and Human Development Psychology, Politcs, Sociology and Education

Violent crime in Romania Estimation, etiology and social prevention

Simona Ionela Mihaiu

PhD, Sociology Department, University of Bucharest simonamihaiu@yahoo.com

Abstract. Violent crime is a social phenomenon which records significant rates in comparison with other forms of manifestations of penal deviation in European space and implicitly in Romania (Eurostat, Crime records by the police, 2012). The repercussions that violent crime have over individuals, the community and social order determines us to place this social phenomenon in a central position of study. If we consider violent crime from the point of view of penal deviation, we propose the emphasis of conceptual and theoretical fundamental analysis perspectives and the presentation of the first results of the corresponding to the PhD. thesis entitled "Violent crime in Romania. Estimation, etiology and social prevention". At this point, our purpose is to build a violent crime analysis perspective in Romania and sound the alarm about costs and social significance of violent crime acts.

Keywords: *criminal deviance, violent crime, social significance.*

1 Introduction

Our PhD. thesis and implicitly the present paper propose the analysis of violent crime as an aggravating form of manifestation of penal deviation. As an extremely complex phenomenon, violent crime has an extremely dangerous character as far as legal norms and social values for individual safety and community order are concerned, implying dysfunctionalities both in the public and private area.

As a response to social reality, specialized literature emphasises the fact that, for many contemporary researchers, violent crime represents a central topic, and the focus is not only on identifying determining factors, dimensions and forms of manifestation, but also on developing some models of optimal intervention against criminal acts as applied to individual and society (see Cusson, 2005; Farrington, 1994; Bandura, 1962; Krug, 2002).

In fact, if we refer to violent crime etiology and wish to understand the interdependence between the determining factors of such a behaviour, it is necessary to analyse the aggravating forms of penal deviation by reporting to macrosocial processes like: the state of anomie specific to society, the rural-urban migration and big cities' overcrowdedness, industrialisation with massive population dislocation and the formation of a community with a low level of social cohesion. By understanding the fact that microsocial factors have an important part in violent crime acts, we consider that it is also necessary to consider aspects like: the low level of education and instruction, drug and alcohol consumption, belonging to a family with penal background, the low development of emotional intelligence, conflictual familial climate.

Starting from these theoretical and applicative reasons, the present paper has the following objectives: conceptual defining of violent crime as a social and antisocial phenomenon, presenting the theoretical orientations that are considered adequate for violent crime etiological and explicative knowledge and shaping the quantitative type (statistical analysis of documents) of evolution of this social phenomenon in Romania.

1.1 Specialised literature

As previously mentioned, violent crime conceptual delimitation is the first step in our paper. Thus, we define the sphere of this notion by considering the following researchers' point of view: M. Cusson, M. Quimet and J. Proulx, according to whom, violent crime includes "forbidden acts by law and liable of penal sanction application" (Cusson and collaborators, 1999). In order to offer a specific definition, so a less general one of the debated concept, we consider P. Bonfilis's thesis, according to which "violent crime can be understood as serious acts that bring direct, intentional and physical touch to the victims" (Cusson and collaborators, 1999).

Taking into account the fact that the paper is centred on violent crime in Romania, we have chosen to correlate the advanced definition of P. Bonfilis with the legislative framework stipulated in the Romanian Penal Code defining the sphere of this social phenomenon as "acts that bring serious physical prejudices directly to the victim" (2010). Therefore, in our analysis, we define the following as being forms of manifestation of violent crime: crimes against people (murders, death causing blows, attempted murder, rape, aggravated mayhem) and crimes with double state (robbery – crimes against people and their heritage).

Considering violent crime conceptual delimitation, we will review the main theoretical orientations selected towards the etiological knowledge of our object of study.

The sociological scenario of violent crime etiology highlights the relationship of interdependence between individual and societal factors that lead to the development of violent criminal behavior, this process occurring as a result of discrepancies we meet inside a given social system, whether we are talking about community, neighbourhood, family or school.

The general hypothesis on which violent crime explicative and etiological approaches are built is that biological, psychological and psychiatric theories cannot fully explain crime and criminal behaviour.

Thus, sociological theories focus on social and psychosocial theories like social learning and imitating criminal behaviour, stigmatisation, social disorganisation or the state of anomie, cultural and environmental influence, negative socialisation; which emphasise the fact that determining such a behaviour is predominantly exogenetic.

As far as we are concerned, we have chosen violent crime from the point of view of the theory of anomie (Durkheim, 1974; Merton, 1938;), the theory of criminal subculture (Cohen, 1955) and the theory of differential association (Sutherland, 1970). Therefore, we consider the ecological model advanced by Krug (2002), which was nevertheless initially debated by biologists and psychologists, according to the Canadian author's model we are offered an integrative perspective which implies considering individual, relational, community and society factors. We have selected the previously mentioned theories because we have considered that they can fully comprise the violent crime sociological specificity scenario, emphasising the interdependence relationship between the individual, the community, the familial and cultural background where he lives.

The psychological perspective of violent crime etiology focuses on the sociological and psychological factors of the individual with those of the environment in which he is socialised. The psychosociological orientations "can identify certain personality characteristics that in certain contexts of social learning and experience, predispose certain individuals towards delinquency acts" (Giddens, 2000). From this point of view, we fundament violent crime psychosocial analysis by making references to learning theories focused on imitation and social learning (see Tarde, 1890; Bandura, 1962; Sutherland, 1966) and the theories of control (see Reckless and collaborators, 1973; Hirschi, 1969; Matza, 1964).

The psychopathological scenario, without necessary excluding the influences of social and biological factors in determining the tendencies towards violent crime and aggressiveness and focuses especially on psychic characteristics, innate or acquired of the individual who manifests a delinquent or criminal behaviour. The general hypothesis which generates psychopathological studies implies the fact that social and biological factors cannot fully explain murder and crime because the real cause of criminal behaviour is in the human psyche, in the flows of structures and functions (see Freud, 1991; Born, 2003; Hare, 1998; Côté, 1999; Butoi, 1996).

From criminological perspective, M. Cusson considers that *delinquency is a chosen way of life*, basing himself in developing his theory on the research results that have emphasised that certain categories of offenders who have committed a large number of crimes represent an important rate among the total crime acts in the particular society.

Although the validity of his considerations is demonstrated only by the category of hyperactive offenders, we consider that the principles advanced by Cusson (2005) can explain a number of violent crime acts, although it cannot explain crimes like murder by a person who has not committed any previous crime. Thus, in completing the criminological scenario, we use theories in order to explain violent crime which is focused on the cultural and socialisation dimension of personality (Fromm, 1941), theories of criminal personality (Pinatel, 1963) and psychomoral theories (De Greef, 1946).

We observe that irrespective of our position in understanding and explaining violent crime, we cannot include this phenomenon in a single theoretical orientation. Be it complementary or contradictory theories, we often reach a multicausal determinism of criminal behaviour. In this context, we consider the fact that it is important to mention the fact that as far as we are concerned, we have chosen a multicausal or a simultaneous theoretical approach of a plurality of factors (individual, social, medical, biological, cultural).

1.2 Methodology

The conceptual and theoretical analyses of violent crime are completed by the developing research field and are to end in December, 2012. This is why we present the proposed methodology in what follows and in the chapter of *Empirical Analysis* we will debate the results of the research first stage, i.e. the analysis of statistical documents issued by European and national organisms.

1.2.1. *The purpose of the research* of our PhD. thesis relies in identifying violent crime phenomenon characteristics and individual and social correlations of violent crime acts in Romania.

For this purpose, we propose the following specific objectives:

- Setting violent crime in the general context of crime in Romania;
- Recording, classifying and analysing violent crime acts in Romania;
- Knowing the ways of subjective reports of authors towards violent crime acts;
- Setting the influence, interdependence, determination reports between the violent crime phenomenon and other social phenomena (the low standard of living and educational level, social mobility, the state of anomie that characterise our country at present and the low level of social cohesion);
- Formulating an etiological theory of violent crime behaviour and of violent crime acts in Romania.

1.2.2. *The research universe* includes at its highest level of generality the authors of crime acts in Romania, in order to assure an optimal framework of analysis. In the second stage of field research, the analysis group will be reduced to 400 jailed individuals for having committed murders, rape, death causing blows, aggravated mayhem, robbery. The last stage of field research consists of applying 40 semi-structured interviews to the convicts in Craiova Penitentiary, Târgsor Penitentiary and Mărgineni Penitentiary, Romania.

1.2.3. These are the *methods and techniques proposed for usage*:

• Document analysis method:

- Studying the official statistics issued in the form of yearbooks, statistical documents or research reports, in order to create a database on crime manifestation in Romania;

- Studying the psychosocial cases of perpetrators of acts of violent crime sent in the three prisons proposed for analysis in order to know their history of criminal behaviour, socio-demographic data and psychological characteristics.

Interview method:

- Semi-structured or semi-directive interview applied to individuals sent to prison and who satisfy the features mentioned in the research universe. By using this work method we want to know the subjective ways of reference to the criminal act (the description of the act, justification, attitudes and feelings).

• Observation method:

- Unstructured observation will be used during the research on detention facilities, its specialists and interviewed imprisoned population.

1.2.4. In correlation with the research objectives, methods and work techniques, we propose the following testing *hypotheses:*

- The etiology of violent crime manifestation is not guarantied by the action of a causal factor, but by the connection of a plurality of complementary and even interferential factors.
- The etiology of crime social contextuality is centred on the individual as main actor. The diversity of individuals' psychical and sociocultural identities implies the fact that these do not promote the same actions and behaviours, even if they are placed in cvasi-identical etiological structures.
- The models of the promoted status-roles within the source families of partners are matrices for models promoted in their own family.
- Criminal behaviour is dominantly rational, as a result of a pragmatic calculation between the act's advantages and assumed risks.
- The etiology of crime social contextuality is filtered by the actors' favoured/disadvantaged psychical structures.

The purpose and the objectives of the PhD. thesis, the proposed testing hypotheses and the working methodology is fundamented by the necessity of completing a specialized literature theoretical and applied order in Romania. At present, although violent crime acts in our country represent an important rate among the total crime acts, specialized analyses and studies covers only partially the etiology problem and this social phenomenon's forms of manifestation, especially if we refer to penal deviations specific approaches. Moreover, we observe the necessity of developing intervention models and strategies that would represent a real support in reducing and preventing violent crimes in present Romanian society, affected by a series of economic, cultural, social and educational problems.

1.3 Empirical analysis (The first field results)

The first stage of research field was made over a 30 days period and it involved statistical data analysis of national and international organisms: Eurostat, Romanian Police General Inspectorate, National Penitentiary Administration, National Institute of Statistics. During the process, we were interested in analysing violent crimes in Romania and the reference period was between 1990 and 2009.

Generally, violent crime in Romania has a special social significance because statistical dates emphasise the fact that during 2009, from a total number of 34.226 of people irrevocably sentenced to prison, 5.801 committed violent crime acts against individuals: murders, death causing blows, attempted murder, aggravated mayhem, rape (Romanian Statistical Annuary, 2010).

1.3.1. Tendencies of the evolution of murder offence rates

Murder offence, of all manifestation forms of violent crimes, reaches the highest rate in Romania during 1990 – 2009, after robbery crime (Romanian Statistical Annuary, 2010).

According to the statistical data analysis of the National Institute of Statistics, we observe the fact that during the first three years after the fall of the communism regime, the number of murders committed in our country has continually increased, reaching a maximum rate in 1992, when 1.556 murders were condemned.

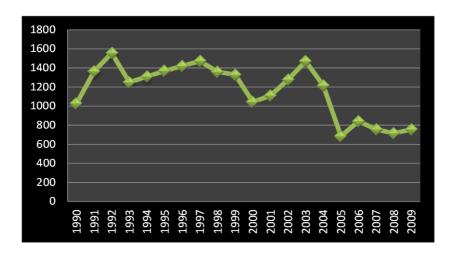


Figure 1: Evolution of murder offences

Source: National Institute of Statistics, *Romanian Statistical Annuary*, Chapter "Justice", 2010

During **1990**-2004, approximately 1.500 murder offences were committed/12 months, and from 2005 these did not reached 1.000 crimes/12 months. In the following years of our reference period, we nevertheless observe new tendencies of increasing the number of murders, reaching 715 acts in 2008, and 756 acts in 2009. By adding all the murder offences during 19 years that we have analysed, according to statistical date from the National Institute of Statistics, 23.332 murders were committed in Romania.

As far as attempted murder crimes are concerned, statistical data from the Romanian Police General Inspectorate points out the fact that between 2004 and 2009, 2.702 cases were inquired. Nevertheless, there are no public information related to the number of attempted murder crimes that were tried and sentenced, and this is the reason why at present, we cannot make a comparative analysis with other violent crime manifestation forms.

After the fall of communism in 1989, Romanian society suffered a series of political, social, economical and cultural transformations and these transformations

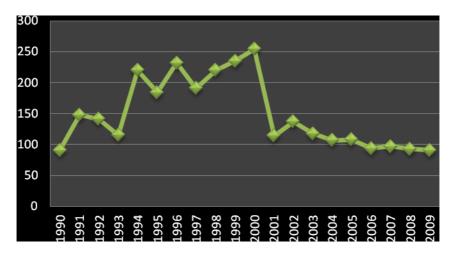
sometimes had dramatic dimensions. There were dysfunctionalities in the state institutions, judicial, moral, familial power and authority were diminished by the sudden change of political regime and interpersonal relations were influenced by new social values and norms.

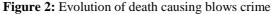
If during the communist period, Romanians had been forced to reduce expenses and needs, after 1989, when political changes happened and the economy started to develop, people wanted to improve their standard of living and became impatient in the process of waiting new laws to be instituted, as old ones had lost their authority. In such situations, individuals lose their social realities "we no longer know what is allowed and what is not permitted, what is fair and what is not fair, legitimate or illegitimate requisitions and expectations" (Durkheim, 1974).

1.3.2. Tendencies of the evolution of death causing blows crime rate

Violent crime manifestation forms, blows or death causing injuries are acts "which produce damage to the victims integrity or health and this damage is followed by the death of the victim" (Romanian Penal Code, 2010). Of all violent crimes, the death causing blows crime rate is rather low, in comparison with the number of murder offences and robberies.

We will continue by presenting the dates from the National Institute of Statistics, between 1990 and 2009:





Source: National Institute of Statistics, *Romanian Statistical Annuary*, Chapter "Justice", 2010

The death causing blows crime, like murder offences, has strong causality relations with the changes underwent by the Romanian society after the fall of communism in 1989. Moreover, Romania is still affected today by a state of social anomie, which accounts for violent crime etiologies. As previously mentioned, we do

not overlook the importance of biological, psychological and psychiatric factors, but it is important to emphasise social causality because although violence and violent crime can be approached and interpreted from various scientific perspectives, in order to obtain a global image, it is necessary to consider the fact that social life is a key factor in individual behaviour.

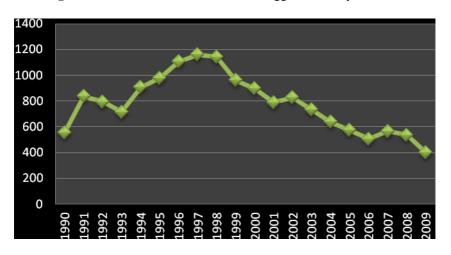
1.3.3. Tendencies of the evolution of aggravated mayhem crime rate

By including aggravated mayhem crime in the violent crime category raises a series of discussions generally based on legislative differences between societies.

According to Romanian legislation, aggravated mayhem is the act which determines prejudicing "body or health integrity, an injury whose recovery requires medical attention for more than 60 days" (Romanian Penal Code, 2010). The category of violent acts circumscribed to aggravated mayhem crime includes those which can imply the loss "of a sense or organ, suspension of these functions, permanent physical or psychical invalidism, deformation, abortion, or endangering a person's life" (Romanian Penal Code, 2010).

Of all mayhem crimes between 1990 and 1996, approximately one fifth represents aggravated mayhem crimes. During 1990-2007 17.539 aggravated mayhem crimes were recorded, an average of 974 crimes per year, according to Ecaterina Balica, in her paper *Violent Crime* (Balica, 2008).

Although the number of aggravated mayhem crimes has dropped considerably, it still remains a social problem which affects personal, medical and social individual lives.





Source: National Institute of Statistics, Romanian Statistical Annuary, Chapter "Justice", 2010

According to the statistical data analysis of the National Institute of Statistics, we observe the fact that aggravated mayhem crimes evolution follows a similar tendency with those of attempted murder crimes and death causing blows.

1.3.4. Tendencies of the evolution of rape offences

The rape offence is one of the most serious forms of interpersonal violence which implies dramatic physical, emotional and psychological consequences for the victim. Romanian legislation is responsible for the necessity to prevent such kind of acts, by punishing any " sexual act, of any nature, with another person, by force or taking advantage of the person's impossibility to defend oneself or to express one's wish" (Romanian Penal Code, 2010).

According to the statistical data analysis of the National Institute of Statistics, we observe the fact that in Romania the number of rape offences has had a similar evolution with other violent crime acts (death causing blows, aggravated mayhem, murders), in the sense that between 1990 and 1997 it reached high rates. From 1998, the number of rape offences began to drop, with a new maximum point in 2003, when 792 violent crime acts of this type were recorded.

We will continue by presenting the statistical data during 1990 and 2009:

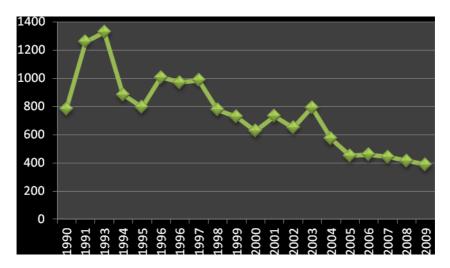


Figure 4: Tendencies of the evolution of rape offences

Source: National Institute of Statistics, Romanian Statistical Annuary, Chapter "Justice", 2010

Sociological theories generally approach rape offences starting from familial violence, alcohol and drug consumption, the low level of education, interpersonal disputes. As far as we are concerned, in order to complete the etiological table of rape offence, we propose that in the second stage of field research we make a comparative analysis starting from the victims' and aggressors' resident background, from the

level of education and type of relationship between the victim and the aggressor. We also consider that resides the etiological aspects of psychological and sociological nature, we need criminological completions (according to De Greef's *The Act Theory*).

1.3.5. Tendencies of the evolution of robbery offence rate

According to Romanian legislation robbery represents "committing a theft by using violence and threats or by inducing a state of unconscious or incapability to defend oneself, as well as theft or use of other similar means in order to preserve the stolen good or remove the traces of the crime or to guarantee the offender's escape" (Romanian Penal Code, 2010).

From the sociological point of view, robbery offence has a double state, a violent act committed both against heritage and person. This is also reflected in the Romanian legislation, which stipulates three aggravating forms of robbery offence:

- a) The first aggravating form exists when: the author is masked, transvestite of disguised, when the act takes place during the night, in a public place or in a means of transport.
- b) The second aggravating form exists when the robbery is committed by two or more people together, when the attacker has a narcotic or paralysing substance, in a house or its belongings, during a calamity.
- c) The third aggravating form exists when the robbery has produced extremely serious consequences or resulted in the death of the victim (Toader, 2007).

Starting with the classification provided by the Romanian lawgiver, we retain for analysis the third aggravating form of robbery offence, i.e. the situation in which the act has produced extremely serious consequences or resulted in the death of the victim.

We will continue by presenting the statistical data during 1990 and 2009:

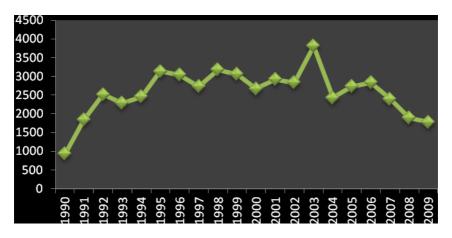


Figure 5: Tendencies of robbery offence

Source: National Institute of Statistics, *Romanian Statistical Annuary*, Chapter "Justice", 2011

According to the statistical data analysis of the National Institute of Statistics, in Romania, robbery offence registered a mild tendency towards a low level in 2009, when 1.780 acts were condemned, in comparison with 2008 when 1.895 cases were condemned. Robbery as a crime against heritage holds a 2 place condemned figure from 2000 after theft offence. In 2009, from the total number of 1.780 irrevocably jail sentences of robbery offence, 11 involved the death of the victim. According to the data of Police General Inspectorate, between 2004 and 2008, there had been 84 robbery offences that involved the death of the victim.

Unlike violent crime acts against a person, committed especially in the rural area, violent crime acts against heritage are committed especially in the urban area. For example, in 2009 5.175 crimes against heritage were committed in the rural area and 8.516 crimes of the same type in the urban area (*Statistics of the main activities of the Romanian Police between 2004 and 2009*).

Although the statistical data emphasises the fact that the number of robbery offences registers a mild tendency towards a low level, it is important to mention that from the point of view of global violent crime, the situation is serious.

This is because we consider that most robbery offences in Romania in the last two decades are closely related to social problems like: permanent change of judicial norms and legislation, the low level of the standard of living, the extremely high rate of unemployment, low efficiency and functionality of state institutions.

Conclusions:

The PhD. thesis briefly presented in this paper approaches an important social phenomenon in the Romanian society, which infringes the right to life, integrity, safety and freedom. If a first analysis of the infringement of these laws stipulated in the Universal Declaration of Human Rights refers to penal law, in fact, violent crime involves a series of socioeconomic, psychological, medical, cultural implications.

Introducing violent crime in etiological study involves in the first stage the correct placing of this social problem within penal deviation, and for this reason, we have considered that it is necessary to remind the definition of the concept of violent crime and its forms of manifestation.

Our analysis reveals the fact that all study domains centred on the violent crime problem focus on psychological, biological, criminological, sociological and psychiatric factors. This is the reason why we have chosen to briefly present several theoretical orientations that we consider appropriate for violent crime etiology.

Moreover, although our approach is mainly sociological, we can reduce or exclude the interdependence of certain factors: the author's and the victim's lifestyle, criminal past of background family, individual characteristics, standard of living, the educational level of the authors and the relationship between the author and the victim.

As far as statistical data is concerned, we observe sensible drops in the number of violent crime acts, especially if the comparative analysis is made between 1990 and 2009. Even so, generally, violent crime holds a high rate of the total number of crimes committed in Romania, and this is why we propose to continue specialised studies and analyses. In this context, we need analyses which would focus on the

authors' and victims' level of education, on the type of relationship between the actors involved, on the process of committing crimes and especially on the subjective representation of the act. Moreover, we cannot ignore pre-offence motives specific to a violent crime behaviour because although these aspects are generally known, we can only make prevention and control strategies of such a social phenomenon based on complex results.

This is why we mention the fact that this research stage is only the first step of a complex analysis, which would involve several etiological aspects of violent crime behaviour. Starting from individual, familial, social mechanism of violent crime is the only way to create models of pertinent intervention that would respond to Romanian's society real needs.

References

Balica, E., (2008), Violent crime. Trends and risks factors, Oscar Print Publishing, Bucharest.

Cusson M., (2005), *Delinquency, a life choice. Between pleasure and crime*, Hurtubise Publishing, HMH Itée, Quebec.

Cusson M., Proulx J., Quimet M., (1999), Criminal violence, University Laval Press, Quebec.

Durkheim E., (1964), The rules of sociological method, Scientific Publishing, Bucharest.

Eurostat (2012) Crime records by the police, Online http://epp.eurostat.ec.europa.eu

Giddens A., (2000), Sociology, BIC ALL Publishing, Bucharest;

National Institute of Statistics, (2010), Romanian Statistical Annuary, Chapter Justice, Online www.insse.ro

Romanian Penal Code, (2010) Hamangiu Publishing.

Police General Inspectorate, (2010), Statistics of the main activities of the Romanian Police between 2004 and 2009, Online www.igpr.ro

Toader, T., (2005), Criminal Law. The special part, Hamangiu Publishing, Bucharest.

The Protection of Vulnerable Groups in Albania

Elda Gjergji, PhD.candidate1

¹ Department of Civic Education, Faculty of Education, University of Elbasan "A.Xhuvani" Email: egjergji@gmail.com

Abstract. In early 1990s, Albania begins to make political and economic reforms toward democracy. Among others, reforms to improve the economy and to protect the rights of its citizens were of the immediate need, and they started to show improvements in Albanian society. However, the economic and social reforms in Albania are accompanied with an increase in the number of people who are considered at risk or outside the mainstream of the society.

The treatment of social protection issues for vulnerable groups in Albania has not been systematically examined despite their vital importance to poverty reduction. In recent years, international development discourse has started to recognize more fully the importance of adequate social protection.

There is no guarantee that social protection and equity issues will come higher up the agenda of the democratization of Albania, not least because the international institutional context in which they are being developed which historically has not prioritized social protection. While civil society involvement has varied substantially, civil actors are often some of the strongest advocates of inclusive and equitable development policies, including adequate social protection in Albania. At the end, this paper is primarily based on documents of state institutions and international organisms working in Albania.

Keywords: *poverty, vulnerable groups, social protection, economic support.*

1 Introduction

In early 1990s, Albania began making political and economic reforms by protecting the rights of man and improving the economy. During this period, the economic and social situation in Albania was very sensitive; from 1990 - 1992 the economic situation continued to weaken, and in 1993 the firs results of privatization were evident with some slight improvements in the economic level. This critical situation of

this period caused the charting of social and political forces to lower the poverty by creating economic help scheme. Nearly one in three Albanians or 917,000 were poor, while 500,000 were considered very poor¹. This poverty is the result of many complex factors going back from the former system of governance adding other causes occurred during the transitional period. According to statistics, 29, 6 % of the population was living in relatively poor conditions, while half of this group was considered extremely poor.

The liberalization and privatization reforms in Albania – one of the most centralized economies of the former communist block, led to the entire transformation of the national production structure. This process had a tremendous impact on the country's economic and social affairs. This complex situation came about partly as a result of the institutional failure. Social protection was among many aspects of socio – economic life where institutions failed to provide a solid frame for transitional reforms. Vulnerable social groups "at risk" of being excluded or fall into the poverty trap could not be properly dealt by social policies and social protection, thus leading to widening gaps between those groups and the rest of society².

Seventeen years from the fall of communism, following the completion of many reforms and improved economic conditions, the need for qualitative social services and protection has become more prominent. Efforts to improve social protection and inclusion are being intensified. This has brought the need for monitoring especially in terms of the link between poverty reduction and social exclusion, so that reducing poverty does not bring more social exclusion and social gaps between different social groups.

Albania has experienced a sustainable pace of economic growth at an average annual rate of around 6%.³ Although reforms have tackled all sectors of economy and education, health and social care, insurance and other protection systems, government structures are still not able to respond adequately to the needs of marginalized and vulnerable groups due to limited financial and human resources as well as inefficient institutional capacities. Therefore, Albania today is facing important challenges on social protection, fight against poverty, inequality and the need to modernize the health and social protection systems based on European standards.

1.1 Informal economy

Informal economy constitutes a serious issue in Albania and the measurement of its size is a real challenge. Albania's informal sector is believed to range between 30 to 60 percent of the nation's GDP and is present in agriculture, construction and microeconomic businesses. Despite the economic growth, Albania is still one of the

¹ "Buletini statistikor" 2011, APNSHS

² Hulme, D., A. Shepherd and K. Moore, (2001) *Chronic Poverty: meaning and analytical frameworks*, Working Paper, Chronic Poverty Research Centre, University of Manchester.

³ Annual INSTAT (2011). Reports. Tirana.

672

poorest countries in Europe. Although the country moved into the group of countries with a high human development index as reported by the UNDP's Global Human Development Report 2007 (68th position with a HDI of 0.801), Albania's GDP per capita stands at 5316 USD (PPP)⁴ which is well below the countries of the region. Unemployment in Albania, though high by western European standards, is lower than in many neighboring countries, (Macedonia 36 percent, Serbia 24 percent, and Bosnia 32 percent). Data are supplied from administrative sources that indicate a sharp increase of unemployment during the early 90s.

1.2. Poverty

Poverty also brings a lack of basic social, political and economic rights for marginalized or isolated social groups. According to statistics⁵, 75 % of poor families deal with other problems in additions to their economic struggles. Poverty is not just a lack of income, it also include a wider base of social needs, including education, living conditions, hygiene and discriminations for this vulnerable segment of the population. In other words, poverty means deprivation of basic human rights.

1.3. Emigration

Emigration has been one of the main characteristics of the economic and social changes that have taken place in Albania since 1990. It is estimated that more than 25 per cent of Albanians live abroad⁶. No other country in East Central Europe has been affected so much by emigration, and in such a short period of time.

One of the most direct effects of migration abroad was the flow of remittances to emigrants' families. One in three households has declared to have someone of its family as an emigrant abroad while more than 65 percent of these households receive remittances⁷. Remittances have had an important influence in the economic life of the country at microeconomic and macroeconomic levels⁸. Since the emigration is relatively among young people, had brought some other social problems such as: departure of unaccompanied children, abandoning school, growth in the number of single parent families, abandoned children and lonely elderly people, etc.

1.4. Urbanization and chaotic movement

⁴ Ibid.

⁵ Ibid.

⁶ King R and Vullnetari J, Migration and development in Albania. Working papers, Sussex Centre for Migration research, 2003.

⁷ LFS 2007: individuals working in their small farms that do not sell their products, but produce them only for their own consumption, are considered as self-employed. www.instat.gov.al.

⁸ King R and Vullnetari J, Migration and development in Albania. Working papers, Sussex Centre for Migration research, 2003.

Urbanization and chaotic movement of the population results in rapid transition, like the Albanian transition to democracy. It has occurred primarily informally through the creation of illegal housing by low – income families. This social phenomenon is accompanied by increasing economic poverty, lack of employment and difficulty of adjusting to urban life raises the risk of young becoming involved in criminal activity, children quitting school because of poverty and limited educational possibilities in the new area, etc.

1.5. Low public security service

Low public security service is another result of this transition period, where the conditions for crime can raise. The vulnerable groups like: children, women and girls and more exposed at crimes. Trafficking of people, in particularly of girl, women and children was a severe social problem. During the past years, Albania has become a place from which human beings are trafficked to other places. There are no accurate statistics for this phenomenon, both national and international institutions reveal that the problem is greater than the official statistics would imply⁹.

2. Poverty and social exclusion- a gross violation of basis human rights

2.1. Human right based approach

Human rights are the basic rights and freedoms to which all humans are considered entitled. These rights represent entitlements of the individual or groups vis-a-vis to the government, as well as responsibilities of the individual and the government authorities¹⁰. Some rights are ascribed "naturally," which means that they are not earned and cannot be denied on the basis of race, creed, ethnicity or gender. These rights are often advanced as legal rights and protected by the rule of law. Various "basic" rights that cannot be violated under any circumstances are set forth in international human rights documents.

While human rights are not always interpreted similarly across societies, these norms nonetheless form a common human rights vocabulary in which the claims of various cultures can be articulated. Having human rights norms in place imposes certain requirements on governments and legitimizes the complaints of individuals in those cases where fundamental rights and freedoms are not respected. Many believe that the protection of human rights is essential to the sustainable achievement of the three agreed global priorities of peace, development and democracy. Respect for human rights has therefore become an integral part of international law and foreign policy.

⁹ "Mid-term country strategy of social services 2003- 2007" Ministry of Labor and social services

¹⁰ Understanding human rights – Manual on Human Rights Education (2003) Human Security Network, Graz, Austria 2003

The specific goal of expanding such rights is to increase safeguards for the dignity of the person.

In order to address human rights violations, we must strive to understand the causes of these breaches. These causes have to do with underdevelopment, economic pressures, social problems and international conditions. Indeed, the roots of repression, discrimination and other denials of human rights stem from deeper and more complex political, social and economic problems.

2.2. Principal of equality and non-discrimination

Human rights apply to all people simply because they are human. But some people, or groups of people, face particular difficulties in realizing their rights because of who they are or what they believe. People are discriminated against on a wide range of discrimination grounds such as their gender, race, and ethnicity, lack of citizenship, sexual orientation, health, property, age or disability. People belonging to these groups have certain common characteristics or are in a situation that have been shown to make these people more vulnerable to discrimination. Many people suffer from discrimination on a number of grounds at the same time. This is often referred to as "multiple discrimination".

Discrimination may directly result from discriminatory legislation or regulations, in which an explicit distinction is made on the basis of ethnic background for example. Legislation or regulations may also indirectly discriminate against particular groups of people. Indirect discrimination may occur when apparently neutral rules and practices have negative effects on a disproportionate number of members of a particular group irrespective of whether or not they meet the requirements of the job. Governments as well as non-state actors such as companies may sustain discriminatory practices even if they do not intend to do so.

Being discriminated is an abuse of a person's dignity and worth. In addition, people who are discriminated against will have increasing difficulty realizing other human rights like access to health care facilities, access to housing, to work, to a fair trial, to redress, etc.

2.3. International legal framework

Article 2 of the UDHR states that everyone is entitled to the rights it proclaims "without distinction of any kind, such as race, color, sex, language, religion, political or other opinion, national or social origin, property, birth or other status". This principle is reflected in the International Covenant on Civil and Political Rights, the International Covenant on Economic, Social and Cultural Rights (ICESCR). In addition to this, United Nations have expanded on the right to freedom from discrimination in the International Convention on the Elimination of all forms of Racial Discrimination (CERD) and the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW)

2.4 Regional legal framework

Based on the experience of contrasting sex discrimination, a consensus emerged in the mid-1990s around the need for the European Community to tackle discrimination on a number of additional grounds. The result of this process was the inclusion of Article 13 in the EC Treaty, to take action to deal with discrimination on a whole new range of grounds, including racial or ethnic origin, religion or belief, age, disability and sexual orientation. This led to the adoption by the Council in 2000 of two directives that have raised significantly the level of protection against discrimination across the EU. The first directive bans direct and indirect discrimination, as well as harassment and instructions to discriminate, on grounds of racial or ethnic origin. It covers employment, training, education, social security, healthcare, housing and access to goods and services. The second directive establishes a general framework for equal treatment in employment and occupation as well as vocational training. It deals with direct and indirect discrimination, as well as harassment and includes important provisions concerning reasonable accommodation, with a view to promoting access of persons with disabilities to employment and training. In July 2008 the European Commission has adopted a proposal for a Council Directive on implementing the principle of equal treatment between persons irrespective of religion or belief, disability, age or sexual orientation outside the field of employment. The proposal covers access to goods and services, social protection, health care and education, but does not cover purely private transactions between individuals.

2.5. Social exclusion

The concept of social exclusion¹¹ is new to post communist Albania. Social exclusion comes as a result of poverty, weak governance, slow decentralization, insufficient social policies, inefficient targeting of poor households as well as inadequate implementation of laws. There are considerable gaps between people living in urban and rural areas, between those living in the north and those in the rest of the country, and particularly between Roma/Egyptians and the rest of the population. According to various studies and policy papers, the most affected groups in Albania are children in need, exploited women, disabled (mentally/or physically) people, unemployed persons; pensioners and elderly people as well as Roma and Egyptians.

¹¹ The Social Inclusion Strategy, approved on January 2008 by the Council of Ministers, is one of the most important government policy document and also an imperative strategic document in Albania's way towards integration to the European Community. In this framework, Social Inclusion is considered as one of the priorities of the current government, with poverty reduction as its main focus, which will be ensured not only through economic development. It focuses on poverty and social exclusion risks that remain even after the onset of economic growth. As a crosscutting strategy, it is fully consistent with the underlying sector strategies and in particular those policies and institutional arrangements described in these strategies that aim to assist vulnerable individuals, families and groups in the community so that they are able to operate on their own, to be self-sustaining and to have the same rights as other members of society.

The calculation of absolute poverty line in Albania is based on the World Bank methodology with the data collected through the LSMS¹². The percentage of population under poverty absolute line has decreased significantly during the period 2002 - 2005. Indeed, the number of people below the national poverty line fell from 25.4% of the population in 2002 to 18.5% in 2005^{13} . This reduction is explained by the sustainable growth rate experienced during this period accompanied by the high level of remittances. The reduction of poverty in absolute terms has been more significant in urban rather than in rural areas.

On the contrary, poverty measurement according to the EU Laeken methodology shows that poverty rates have increased between 2002 and 2005. However, there is a substantial difference between the poverty situation in the rural and urban areas.

After 2005, social policies documents have tried to target all vulnerable groups in view of EU objectives on poverty and social exclusion and taking into consideration the Laeken indicators as a tool on describing and monitoring poverty reduction and exclusion¹⁴. Laeken indicators are partially used but still in these documents poverty measurement and its assessment refer to the absolute poverty line. This is the reason why it is possible to observe a significant reduction of poverty in the country. However, relative poverty definition shows an increase of poverty.

3. Empowering the Vulnerable groups

3.1. Women

The criminal code penalizes rape, including spousal rape; however, victims rarely reported spousal abuse, and officials did not prosecute spousal rape in practice. The concept of spousal rape was not well established, and authorities and the public often did not consider it a crime. The law imposes penalties for rape and assault depending on the age of the victim. For rape of an adult, the prison term is three to 10 years; for rape of an adolescent between the ages of 14 and 18, the term is five to 15 years and, for rape of a child under the age of 14, the sentence is seven to 15 years.

Domestic violence against women, including spousal abuse, remained a serious problem. During the year police reported cases of domestic violence and the government pressed charges in cases. The Department of Equal Opportunities at the Ministry of Labor, Social Affairs, and Equal Opportunity covers women's issues, including domestic violence.

The government did not fund specific programs to combat domestic violence or assist victims, although nonprofit organizations provided assistance. NGOs reported that an

¹² LSMS – Living Standards Measurement Study of World Bank

¹³ Institute for Contemporary Studies (ISB) "Social Inclusion and social protection in Albania", European Communities, 2008

¹⁴ Social Inclusion Crosscutting Strategy 2007–2013 of MoLSAEO, approved on 2 February 2008.

estimated eight domestic violence hotlines operated. The hotlines, serving mainly the northern part of the country, each received approximately 25 calls per month from women reporting some form of violence. NGOs operated four shelters for battered women in Tirana, Vlora, Elbasan, and Gjirokaster. During the year NGOs and police noted a substantial increase in reports of domestic violence, primarily due to increased awareness of services and more trust in the police.

According to government figures, there were 1,744 cases of domestic violence reported during the year, compared with 1,063 in 2009. Often the police do not have the training or capacity to deal with domestic violence cases.

In many communities, particularly those in the northeast, women were subjected to societal discrimination as a result of traditional social norms that considered women to be subordinate to men.

Reproductive rights are generally respected by the government. Couples and individuals have the right to decide freely the number, spacing, and timing of their children and have the information and means to do so free from discrimination, coercion, and violence. Citizens have access to contraception. Under the law, health care is provided to all citizens; however, the quality of and access to care, including obstetric and postpartum care, was not satisfactory, especially in the remote rural areas. According to 2008 UN estimates, the maternal mortality rate in Albania is 31 deaths per 100,000 live births. Women are equally diagnosed and treated for sexually transmitted infections, including HIV.

The law provides equal rights for men and women under family law, property law, and in the judicial system. Neither the law nor practice excluded women from any occupation; however, they were not well represented at the highest levels of their fields. The law mandates equal pay for equal work; however, the government and employers did not fully implement this provision.

3.2. Children

The Albanian parliament has approved the law "For the Protection of the Rights of the Child." This law provides the legal and institutional framework for protecting children's rights.

In general parents must register their children in the same community where they are registered. However, according to the Children's Rights Center of Albania (CRCA), children born to internal migrants or those returning from abroad frequently had no birth certificates or other legal documentation and, as a result, were unable to attend school. This is a particular problem for Romani families as well, who often marry young and fail to register their children.

The law provides for nine years of free education and authorizes private schools. School attendance is mandatory through the ninth grade or until age 16, whichever comes first; however, in practice many children left school earlier than the law allowed to work with their families, particularly in rural areas. Parents must purchase supplies, books, uniforms, and space heaters for some classrooms, which were

prohibitively expensive for many families, particularly Roma and other minorities. Many families also cited these costs as a reason for not sending girls to school.

As in previous years, child abuse, including sexual abuse, occurred occasionally, although victims rarely reported it. In some cases children under the age of 18 engaged in prostitution. The penalties for the commercial sexual exploitation of children range from fines to 15 years' imprisonment. The country has a statutory rape law and the minimum age of consensual sex is 14. The penalty for statutory rape of a child under the age of 14 is a prison term of five to 15 years. The law prohibits making or distributing child pornography, and the penalties are a fine of one to five million leks (\$10,000 to \$50,000) and a prison sentence of one to five years. Child marriage remained a problem in many Romani families and typically occurred when children were 13 or 14 years old.

3.3. Persons with Disabilities

The constitution and law prohibit discrimination against persons with physical, sensory, intellectual, and mental disabilities; however, employers, schools, health care providers, and providers of other state services sometimes discriminated against persons with disabilities. The law mandates that new public buildings be accessible to persons with disabilities, but the government only sporadically enforced the law. Widespread poverty, unregulated working conditions, and poor medical care posed significant problems for many persons with disabilities.

During the year the ombudsman continued to inspect mental health institutions and found that while physical conditions in facilities in Vlora and Shkoder had improved, they were not in compliance with standards and remained understaffed.

Inspections of the Tirana Psychiatric Hospital found that specific windows and doors needed to be replaced for safety reasons. The ombudsman regularly conducts inspections throughout Albania and recommended a major legal, organizational, and budgetary review of the country's mental health care system. The admission and release of patients at mental health institutions was a problem due to lack of sufficient financial resources to provide adequate psychiatric evaluations.

3.4. National/Racial/Ethnic Minorities

As visible minorities, members of the Romani and Balkan-Egyptian communities suffered significant societal abuse and discrimination.

The law permits official minority status for national groups and separately for ethno linguistic groups. The government defined Greeks, Macedonians, and Montenegrins as national groups; Greeks constituted the largest of these. The law defined Aromanians (Vlachs) and Roma as ethno linguistic minority groups.

In 2005 the Council of Ministers approved the National Action Plan for the Roma and Egyptian Involvement Decade for 2010-15. The total budget for implementing the five-year plan was expected to be nearly 2.5 billion leks (\$23 million)¹⁵.

The ethnic Greek minority pursued grievances with the government regarding electoral zones, Greek-language education, property rights, and government documents. Minority leaders cited the government's unwillingness to recognize ethnic Greek towns outside communist-era "minority zones"; to utilize Greek in official documents and on public signs in ethnic Greek areas; to ascertain the size of the ethnic Greek population; or to include a higher number of ethnic Greeks in public administration.

During the year government prosecutors continued to appeal the dismissal of charges against Mr. Bollano, the ethnic Greek mayor of Himara, who was found guilty of abuse of office, but whose conviction was overturned on appeal in June 2009. The court convicted Bollano of destruction of government property after he ordered the removal of several new road signs in the Himara district because they were written in Albanian and English but not Greek.

3.5. Societal Abuses, Discrimination, and Acts of Violence Based on Sexual Orientation and Gender Identity

There are no laws criminalizing sexual orientation, and the law does not differentiate between types of sexual relationships. There were lesbian, gay, bisexual, and transgender (LGBT) organizations in the country. The groups operated without interference from police or other state actors. There were reports that individuals were beaten, fired from their employment, or subjected to discrimination due to their sexual orientation. Often these cases went unreported. NGOs claimed that police routinely harassed LGBT persons and transgender sex workers.

4. Social protection policies

4.1. Identification of the contributing factors to increased vulnerability of individuals or groups

Poverty is one of the main contributing factors to increased vulnerability of becoming a human right violation victim. Nevertheless, a multidimensional approach should be considered when assessing the poverty factor. The factors can be divided in the levels of economy, education, ethnicity, and on social level. However, poverty remains one of the crucial factors which is in tight correlation to the other mentioned factors. Answers show that the reasons and contributing factors to increased risks are as follows,: low educational level, bad economic situation, juvenile marriages,

¹⁵ Annual State Budget, 2009

unemployment, lack of parental care, unawareness of the human rights, war conflicts, etc.

The big challenge for an adequate social protection system is the financing, the funding sources of social protection system. Indicators on the efficiency of the system on allocating funds are analyzed in the view of the objective of implementing a social policy that complies with national and EU objectives of social inclusion and protection. From these analyzes, the need for cooperation between different institutions of the system, and different actors emerges as the firs conclusion. In order to support government resources in both local and regional lever has to be forged with a close partnership with NGOs and business communities.

The extension of the social protection system with family and child benefits, as part of the social assistance or social insurance system should be put forward as a recommendation. The social intervention on children will have effects on family poverty, access in female employment, education and health.

Also cash benefits should be integrated with community programs/services for children where the role of social workers and psychologists to be the main one, instead of that of nurses and educators. Educational and recreational activities for children will facilitate the engagement of mothers in the economic and social life.

On the other hand, there is a need for new schemes for protecting family and children in particular. In the context of economic and social problems, traditional values and roles in the family, isolation and lack of proper education/ health and other services, as well as demographic tendencies of reducing birth rates are asking for intervention on family support, on parenting supporting and early child development¹⁶. The fact that women are increasingly employed in the informal sector, or in unpaid family-based activities, women's unpaid care responsibilities on children and elderly are asking for new forms of social protection schemes.

Reforms to tackle informal economy constitute now an important challenge for Albania. Education and the quality of education system in all levels is the element for combating poverty, empowering women, protecting children from hazardous and exploitative labor and sexual exploitation¹⁷. In order to cope with the impact of migration from rural to urban areas there is an immediate need to build new education infrastructure facilities in ratio with the actual trends of migration.

It is important to ensure the participation and integration of people with disabilities through the implementation of legislation provisions that guarantee the application of their rights on employment, easy access to public services and public infrastructure. Implementation of the national Strategy for People with Disabilities and individual programs aiming at the social integration of people with disabilities, most preferably with the active participation of their associations should be promoted. The government protects people with disabilities through subsiding policies, especially for medical

¹⁶ Institute for Contemporary Studies (ISB) "Social Inclusion and social protection in Albania", European Communities, 2008

¹⁷ OECD (2007) Reviews of National Policies for Education: South Eastern Europe: Vol. 1: Albania, Bosnia-Herzegovina, Bulgaria, Croatia, Kosovo. Paris.

services and public transportation. However, it is necessary to allocate the appropriate budget for the implementation of measures/initiative defined by the national strategy.

People living in rural areas, unemployed people living with and older people face a higher poverty risk than the general population. Reforms have to aim at achieving adequate and a sustainable health insurance and social system. This requires a strategy to raise employment and reform the social protection systems.

4.2. The definition of Social Inclusion in Albania

The Social Inclusion Strategy, approved on January 2008 by the Council of Ministers, is one of the most important government policy document and also an imperative strategic document in Albania's way towards integration to the European Community. This strategy remarks a qualitative time in the government social policies and a very important step in crossing towards integrated, preventive and active social policies. It aims our investment and commitment in the social inclusion widening instead of the struggle against social exclusion.

In this framework, Social Inclusion is considered as one of the priorities of the current government, with poverty reduction as its main focus, which will be ensured not only through economic development¹⁸. It focuses on poverty and social exclusion risks that remain even after the onset of economic growth. As a crosscutting strategy, it is fully consistent with the underlying sector strategies and in particular those policies and institutional arrangements described in these strategies that aim to assist vulnerable individuals, families and groups in the community so that they are able to operate on their own, to be self-sustaining and to have the same rights as other members of society.

Conclusion

The transition period accompanied with critical social and economic situation has influenced planning for determining social policies and decreasing poverty. What remains crucial for Albania is the fact that there is a lack of specific studies concerning vulnerable groups. Referring to World Bank studies, the Albanian government has identified a number of groups called "at risk": children, women, youth, elderly people and people with disabilities.

Almost two decades after democracy and privatization were supposed to deliver undreamed-of advances; life in today's Albania is marked by massive unemployment and disillusionment. So it is not surprising that disappointment and low expectations pervade conversations with many men and women. At the beginning of transition, a legal framework was established to provide for the possibilities of structural changes and state property privatization. However, the legal framework itself wasn't sufficient

¹⁸ Albanian Institute for Public Policies (December 2006) *Një sistem monitorimi për Strategjitë Rajonale.*

to respond to the fast political and economic changes that were brought about by the re-emergence of the private sector and capitalism in Albania.

The term 'social security' is hardly ever used either in the Albanian literature on social protection or in the relevant legislation. This is mainly due to the absence of any social right to social security, guaranteed by the Albanian constitution as well as to the predominant socio-political objectives that relate the scope of the social protection to the coverage of working people and needy persons and not to the coverage of the whole population.

The state is the main financial and institutional factor in dealing with these needs which are in the process of changing according to different needs. From this point of view, the existing social services are not able to meet the needs for these services. Tackling social problems through proper policies and institutions is being ranked high at the national political agenda, as a result of many factors related to the country's stage of transition and economic development as well as Albania's efforts to get closer to the EU accession.

References

Albanian Institute for Public Policies (December 2006) Një sistem monitorimi për Strategjitë Rajonale, pp. 39-41.

APNSHS (2011) Buletini statistikor. Tirana.

Annual INSTAT (2011). Reports. Tirana.

ETF Report (January 2008) Migration patterns and human resources development issues in Albania.

Hulme, D., A. Shepherd and K. Moore, (2001) *Chronic Poverty: meaning and analytical frameworks*, Working Paper, Chronic Poverty Research Centre, University of Manchester.

INSTAT (2008) Shqipëria: Trendi i Varfërisë 2005–2008.

IMF (2004) Albania: Selected Issues in Tax Policy Design. A technical assistance report.

Institute for Contemporary Studies (ISB) (2008). "Social Inclusion and social protection in Albania", European Communities, Tirana.

Living Standard Measurement Survey, (2008) Draft, INSTAT and World Bank

OECD (2007) Reviews of National Policies for Education: South Eastern Europe: Vol. 1: Albania, Bosnia-Herzegovina, Bulgaria, Croatia, Kosovo. Paris.

King R and Vullnetari J, (2003). Migration and development in Albania. Working papers, Sussex Centre for Migration research.

Mackintosh, M., 1(999) 'Do healthcare systems contribute to inequalities' in D. Leon and G. Walt (eds), *Poverty, Inequality and Health: an international perspective*, Oxford University Press, Oxford

The Ombudsman, OSCE (September 2003) *Strategjia Kombëtare për përmirësimin e kushteve të jetesës së minoritetit Rom.*

SAMPSON, Steven (1996) All things are possible, nothing is certain: the horizons of transition in a Romanian village." In: D. Kideckel: Local Communities in Eastern Europe. Westview Press.

Understanding human rights – Manual on Human Rights Education (2003) Human Security Network, Graz, Austria

World Bank, (2007) Albania, Urban Increase, Migration and poverty reduction. Poverty Central Asia Region. Report no.40071, AL.June

Normative acts (national and international) and policies

Law No. 8417 of 21/10/1998 "The Constitution of the Republic of Albania" published in the Official Journal, no. 28, 1998 page 1073.

UN Convention on the Rights of the Child, ratified by Law No. 7531 of 11/12/1991 "On the ratification of the Convention on the Rights of the Child".

Law No. 8387 of 30/07/1998 "On some amendments to Law no. 7952 of 21/06/1995".

Law No. 8950 of 10/10/2002 "On the registrar's office" published in the Official Journal, no. 71, 2002 page 1975.

Law No. 9355 of 10/03/2005 "On social aid and services" published in the Official Journal, no. 22, 2005 page 829.

Law No. 3766 of 1712/1963 "On health care" (as amended) published in the Official Journal, no. 16, 1963 page 312.

Law No. 9669 of 18/12/2006 "On measures against domestic violence" published in the Official Journal, no. 150, 2006 page 6113.

Decision of the Council of Ministers No. 368 of 31/05/2005 "The National Strategy for Children and its Action Plan 2005 –2010".

Decision of the Council of Ministers No. 782 of 16/01/2006 "The National Strategy for Youth and its Action Plan 2007–2013".

National Strategy on Gender Equality and domestic violence 2007–2010 of MoLSAEO (2007).

Social Inclusion Crosscutting Strategy 2007–2013 of MoLSAEO, approved on 2 February 2008.

Sectoral Strategy on Social Protection 2008–2012 of MoLSAEO, approved in January 2008.

Annual State Budget, 2007, 2008, 2009, 2010

Women's Informal Employment as an Engine of Poverty Reduction

Souad Belhorma

Center for Studies and Research on Women (English Department), Sidi Mohamed Ben Abdullah University, Belhormasouad23@gmail.com

Abstract. This paper investigates the relationship between women's participation in informal sector and poverty reduction in the city of Fez and, in turn, characterizes ways that those links help into moving women out from social exclusion. Based on critical ethnography, the study uses the investigatory phase to confirm or to refute the theory that women's informal employment is a means to reduce household poverty. The investigation suggests using methodological pluralism by combining qualitative and quantitative methods to get a deep knowledge of women's interpersonal and economic motivations to perform informal activities, as well as the way their income contribute to reducing poverty taking into account the respondents' perceptions and attitudes. Simple random techniques are used as criteria for sample selection. Descriptive statistics and perceptions analysis are used to analyze data. Although the majority of women have vulnerable statuses both at home and at work, their income generation alone may not booster their socio-economic situation in relation to the Informal sector. In this regard, informal women workers can be economically empowered and move out from social exclusion if they cherish equal rights, that could have an improving impact on their power representation in the household and at work. This means that regulating women's work would have a magnificent influence on their lives. This would happen if the organization of work is combined with provision of equal resources and rights as well as equal sharing of benefits.

Keywords: Women, Informal sector, Self-employed, Poverty reduction, Fez.

1 Introduction

The present paper provides some basic information about the Ph.D. project "Women of the Shadows in Focus: Analyzing the Participation of Women in Informal Segments of Work and Poverty Reduction, Fez as a Case Study". Its aim is to investigate the relationship between women's work in the informal sector and reduction of household poverty and, in turn, characterizes ways that those links help into moving women out from social exclusion. This paper is based on desk research demonstrating how I aim to carry the research out in the next phases of the Ph.D. work plan.

II- Review of the Literature

Interest in the informal economy has gained much attention since its discovery in the early seventies. It continues to trigger research to prove useful as a field of interest and hotly debates. The following are four schools of thought clarifying the informal economy in developing countries.

2.1 The Dualist School

This school sees the informal sector as comprised of profitable, marginal and peripheral activities with few, if not to say, with absent links with the formal sector (ILO, 1972). They provide poor people with employment opportunities and income to be solved from times of crisis. (Hart, 1973; ILO, 1972).

2.2 The Structuralist School

The structuralist school expanded the debate about the informal economy to include small, decentralized and more flexible units.¹ They represent subordinate categories to widely spread capitalist firms, serving to reduce their input and labor costs and, in that way their competitiveness with large firms is subject to increase (Moser 1978, Castells and Portes 1989).

3.2 The Legalist School

The legalist school sees the informal sector as comprised of units and entrepreneurs who choose to work informally in order to avoid the costs of registration and taxation, to avoid spending time and efforts in formally regulating their work and converting it into legally recognized units. (De Soto 1989, 2000)

4.2 The Voluntarist School

It regards the informal sector as consisting of micro-entrepreneurs who voluntary prefer to work informally. They operate informally to evade costs of formality such as paying commercial regulations, electricity and rental fees. This means that informal workers deliberately exclude themselves from state's regulations and rights protection (Maloney, 2004).

III- Proposed Research Methodology

The present study is a field survey research using the mixed methods approach to generate credible findings. Data collection will make use of quantitative and qualitative methods. The former will be based on questionnaires, household drop-off surveys, handed to the respondents personally; thereby there is a possibility to write the answers on their behalf since it is expected that many women would be illiterate.

¹ It was popularized by Caroline Moser, Manuel Castells and Alexandro Portes in the late 1970s and early 1980s.

The latter will be based on face to face interviews to enrich the research by stories of women's experiences, behaviors, opinions about their work in the informal sector and how do they manage to reduce poverty, therefore to move out from social exclusion. Moreover, in depth and semi structured interviews will be conducted with experts in the field, scholars, ONG and cooperative members, representatives of Ministry of Employment and Professional Training as well as others from the Ministry of Development, Family and Solidarity in order to gather information about the subject of research.

The field survey is a "one shot survey" based on a sampling model. A group of women informal self-employed in Fez, at a set point in time, will be selected as a representative sample. Moreover, the listing of the members of the accessible population (sample frame) will be based on a snowball sampling of 300 women self-employed. It starts by identifying women who have access to the four cooperatives and the two associations which accepted to work with me during the research period. The selected women must meet the criteria for inclusion in the study. It should be clear here that the reason to choose this type of sampling is due to the inaccessibility of women working in the informal sector due to the nature of their activities, which in most cases are practiced indoors and they are not recognized as work.

Conclusion

As a way of conclusion, although majority of women have vulnerable statuses both at home and at work, it is expected that their income generation from informal activities may not booster their socio-economic situation. Informal women workers can likely be economically empowered and moved out from social exclusion if they cherish equal rights. This situation would have an improving impact on their power representation in the household and at work. Probably, the regulation of women's informal work would have a magnificent influence on their lives. The latter would happen if the organization of work is combined with provision of equal resources and rights as well as equal sharing of benefits. The present study aims to contribute to the academic research and the literature on informality, gender and poverty in Morocco. It should be taken into consideration that the research uses a limited sample. Yet, the major concern is to come out with findings which may be useful to policy makers and stakeholders to help them take action towards women in the informal sector who are deliberately invisible.

References

Castells, M & Portes, A (1989) World Underneath: The Origins, Dynamics, and Effects of the Informal Economy, in Alejandro P, et al (eds), The Informal Economy: Studies in Advanced and Less Advanced Developed, Johns Hopkins University Press, Baltimore, pp.11-37.

De Soto, H. (1989) the Other Path: The Economic Answer to Terrorism, Harper Collins, New York.

----- (2000) The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else, Basic Books, New York.

Hart, K (1973) Informal Income Opportunities and Urban Employment in Ghana. Journal of Modern African Studies, Vol.11, No. 1, pp. 62-89.

International Labor Office, (1972) Employment, Incomes and Equality: A Strategy for Increasing Productive Employment in Kenya. International Labour Office, Geneva.

Maloney, W (2004) Informality Revisited. World Development, Vol.32, No.7, pp.159-1178.

Moser, C (1978) Informal Sector or Petty Commodity Production: Dualism or Independence in Urban Development. World Development, Vol. 6, No.9, pp. 1041–1064.

http://www.socialresearchmethods.net/kb/datatype.php (Accessed 1/11/2011)

Power relations between women and men in Roma communities

Andreea Bragă¹

¹Beneficiary of the project "Constructing and implementing an interdisciplinary innovative doctoral programme concerning Roma issues", co-funded by the European Union through the European Social Fund, Sectorial Operational Programme Human Resources Development 2007-2013, National School of Political and Administrative Sciences, Bucharest, Romania. <u>andreea braga15@yahoo.com</u>

Abstract. The aim of the paper is to investigate the power relations within Roma communities, with a focus on gender relations, and to analyze their impact on violence against Roma women. The research proposal aims to develop an accurate description of who has power over whom in Roma communities, the means through which this power is being exercised in everyday life, and what the mechanisms could be to remedy these inequalities. The proposed study of the power relations and of violence against Roma women will be an empirical inquiry that investigates this acute contemporary phenomenon within their real-life context by using a conceptual framework based on radical feminist theories.

Keywords: power relations, violence against women, radical feminism, Roma communities

1. Introduction

The lack of scientific research regarding different types of violence against Roma women indicates both the marginalization of the phenomenon by the majority and by the Roma minority itself, as well as the difficulty of finding an adequate approach by the academic community and the policy-makers altogether. Moreover, although violence against Roma women is an enduring theme, studies and interventions that deal with it are scarce, especially in Romania. This is why another focus of the study is to criticize and extend the paradigm of radical feminist theories and to develop an adequate political theory in order to comprehend power relations in Roma communities, especially those regarding gender.

Violence is rooted in the dynamics of gender relations and the social construction of femininity and masculinity creates power asymmetry between women and men and determines the role and behavior of individuals. The lack of access to power and the difference between women's experiences and those of men involves a different perspective on power. According to Elshtain (1992), the "power over" and "the power to" represent complementary forms of power and they can be associated to masculinity (formal) and femininity (informal) (Miroiu, 2004). Starting from these theoretical premises, violence against Roma women is a promising perspective both in what regards enriching the academic sphere of feminism and ethnicity, as well as with respect to direct actions that can be developed against violence but that need to be rooted in the realities of the specific gender relations within the Roma community.

The unbalanced relations of power between women and men are expressed both structurally and in terms of behavior. The additional power asymmetry supported by the culture of domination creates a context which limits the access of sensitive subjects in the public attention in order not to reinforce negative stereotypes and to attribute phenomenon like violence against women to an ethnic minority (Schwartzman, 1989; Crenshaw, 1993). In the public life as well as in private relationships, the gender dimension is the main factor leading to an increase in the vulnerability of Roma women in all aspects of life. The reduced visibility of Roma women in the public sphere involves an interpretation of their problems from a biased point of view affected by stereotypes.

There are no previous data regarding the relations of power between women and men in Roma communities and how these relations can result in different forms of violence against Roma women. With a few exceptions, the official data from Romania regarding violence against women do not take into consideration the gender and ethnic dimensions and the process of data collection exclude various types of violence, which minimize the entire phenomenon (Baluță, Bragă, Iancu, 2011). The Roma women's experiences of violence were not a subject of interest either for the feminist or antiracist analyses from Romania.

2. Literature Review

It was not until the 1970s that introducing violence against women on the public agenda and denouncing it as a form of control and oppression contributed to further analysis regarding the relations of power between women and men (Brownmiller, 1975; Dworkin, 1981; Russel, 1975). Studies about women's experiences created the context to conceptualize different forms of violence like sexual abuse, harassment, rape and psychological, social or economical violence (Kelly, 1988;). Although scientific research regarding violence against women escalated, analyses regarding violence against minority women have only been brought to attention in the last two decades (Adelman, Erez, Shaloub-Kevorkian, 2003; Marsh, 1993). Schwartzman (2006) suggests that the government is not the only power that is capable to deprive people of their rights and, under conditions of racial and gender hierarchy, these phenomena are intertwined and their effects are compounded; she takes into consideration the context of male power for issues like domestic violence, rape, abortion, etc. According to MacKinnon, "men's forms of dominance over women have been accomplished socially as well as economically, prior to the operation of law, without express state acts, often in intimate contexts, as everyday life" (MacKinnon, 1989, p. 161). The distribution of power and resources between Roma women and men within the private sphere could give instances of different types of gender oppression. In this context, it is important to interpret human rights against different types of oppression, including systems of power like individuals and groups (Schwartzman, 2006, p.33).

3. Proposed Methodology

The present research will contribute to the understanding of the relations of power between women and men in Roma communities through an innovative conceptual and methodological framework for the research community. In Romania, there is no in depth study regarding the relations of power in Roma community and the Roma women's experiences of different forms of violence.

The present research is base on the conceptualization and definition of relations of power in Roma communities and the interconnection with the violence against Roma women. To investigate the relations of power in Roma communities and to analyze how these can influence violence against Roma women I will design in depth unstructured interviews in order to answer the following research questions:

- 1. Which are the structures of power between women and men in Roma communities?
- 2. Does the structure of power between Roma women and men reinforce gender inequalities and violence against women?
- 3. Does the social context influence the structures of power between Roma women and men?
- 4. Is gender an important factor for one's social power and position in Roma communities?
- 5. Structures of power in the Roma community Who has power over whom in terms of rights, equality, liberty, autonomy?

4. Conclusions

The research proposal is the first of its kind in Romania and would create the necessary climate for dialogue between European researchers concerning the influence of the relations of power on different form of violence against Roma. This subject has not been researched enough at a national level so this paper should prove to be important in identifying problems and proposing solutions regarding this issue. By taking into consideration Roma women's interests and needs through experientially sensitive researchⁱ theories should develop out of women's experiences (Valentine; 1989) and try to achieve positive change in Roma women's lives. The relevance of this project is given by the analysis of the relations of power in Roma communities and by identifying strategies in accessing power and mechanism of empowerment at the local level.

5. References

Adelman, M., Erez, E., Shalhoub-Kevorkian, N., 2003, Policing Violence Against Minority Women in Multicultural Societies: "Community" and the Politics of Exclusion, *Police& Society*, Nr. 7, pages 105-133.

Băluță, O., Bragă, A., Iancu, A., in press, *Impactul crizei economice asupra femeilor*, UNDP, Maiko, Bucharest, în curs de publicare, page 77.

Brownmiller, S., 1975, Against our Will. Men, Women and Rape, Bantam Book, New York.

Dustin, M.; Phillips, A., 2008, Whose agenda is it?: abuses of women and abuses of `culture' in Britain, *Ethnicities*, Vol. 3, No. 8, pages 405-424.

Kelly, L., 1988, Surviving Sexual Violence. Cambridge, Polity Press

MacKinnon, C., 1989, Sexuality, pornography, and method: Pleasure under patriarchy, *Ethics*, Vol. 99, pages 314-436.

Marsh, C.E., 1993, Sexual assault and domestic violence in the African American community, Western Journal of Black Studies, No.17.

Elshtain, J.B., 1992, The power and the powerlessness of women, in Miroiu, M., 2004, *Drumul catre Autonomie*. *Teorii Politice Feministe*, Polirom, Iași, page 179.

Russell, D.E.H., 1975, *The Politics of Rape: The Victim's Perspective*. New York, Stein and Day.

Schwartzman, L., 2006, *Challenging liberalism: Feminism as political critique*, University Park, Pennsylvania State University Press.

ⁱ The concept involves adopting sensitive people-centred techniques (in depth unstructured interviews, participant observation, focus groups), Valentine Gill,1989, *Women's fear of male violence in public space: A spatial expression of patriachy*, Thesis submitted for the degree of PhD, Department of Geography, University of Reading, page 20.

A survey of intention to leave among nurses in the Hail region's hospitals in Saudi Arabia: the role of job satisfaction, job stress and burnout.

¹Salman Alsaqri

¹PhD Candidate, BSN, MSN, RN (Health Science College) RMIT University, Ksa_77@yahoo.com

Abstract

Attracting people to nursing is reported to be difficult. Retaining nurses within their existing position is a global issue that is particularly difficult in specialty areas. Understanding the factors that affect nurses' job satisfaction is important in developing strategies for mitigating the nursing shortage. Additionally, this study will examine the job satisfaction, job stress, burnout and intent to leave and the organisational climate factors that are associated with the nurses' preference whether to stay with an organisation or to leave. Job satisfaction of nurses is essential to decrease the severe nursing shortage. This topic of research is designed to contribute to the future knowledge base on nursing burnout factors and enhance awareness of these factors that is paramount to health care systems today. This study is also important because findings may assist nursing administration and hospital administration in implementing efforts and strategies that can decrease burnout and related factors of job stress, job satisfaction, and intention to leave. Findings from this research may also enhance further understanding of the phenomenon of turnover intention within the profession of nursing in the Middle East in general and in Saudi Arabia in particular. Finally, an understanding of the factors affecting nurses' decision to leave the job could assist employers to develop responsive strategies to address nurse recruitment and retention in Saudi Arabia. This study will use a quantitative exploratory descriptive design to identify and explore the relation between job satisfaction, job stress, burnout and intention to leave in a group of hospitals in the Hail region in Saudi Arabia.

Keywords: Intention to leave, Nursing stress, Job satisfaction, Burnout, Saudi Arabia

1 Introduction

Nurses are the largest group of healthcare professionals and the quality of health care is highly dependent on an adequate supply of qualified nursing personnel (Canadian Nursing Advisory Committee, 2002; Hassmiller & Cozine, 2006). Studies show that there is a direct relationship between the availability of nursing staff and nurses' productivity and, hence, the quality of care (Ceci & McIntyre, 2001; Aiken et al., 2002; Dehghan-Nayeri et al., 2006). A nursing staff shortage is one of the

challenges that nursing has faced previously (Abualrub, 2007). Nowadays, it has become a really challenging issue for the healthcare systems all over the world (Hirschfeld, 2009). Such a fact is completely clear not only to nurses, but also to other healthcare providers and hospital administrators. Aiken et al. (2001) stated that nurses are the core of the healthcare system; without this core, the cell will not survive. This is exactly what will happen to the healthcare system if the number of nurses keeps diminishing. It is also believed that resolving the nursing shortage is one of the most critical steps in providing better health care for people living in developing countries (Buchan, 2002; Nikbakht Nasrabadi & Emami, 2006). Abualrub (2007) stated that the shortage of nurses is a limiting factor in the development of an optimum system of healthcare delivery in the Middle East. Furthermore, studies show that one of the most important factors regarding the nursing shortage is job satisfaction; that is, the greater the level of job satisfaction, the smaller the nursing shortage (Zangaro & Soken, 2007). The aims of this study are To identify factors that may influence or contribute to the retention of registered nurses, To ascertain the nurses, job stress, level of job satisfaction and burnout, To investigate the nurses' intention to leave an organization. To make recommendations at both departmental and organisational levels to enhance the retention of nurses and To influence planning, policymaking and professional development in relation to the retention of registered nurses.

2 Literature Review

Patient care is provided by nurses who are often working in hospitals. Therefore, if hospitals are worried about their consumers' satisfaction, they need to learn how to retain an adequate number of nurses and how to improve the nurses' job satisfaction. The nursing shortage is not only affecting nurses. Services have been reduced, surgeries cancelled and units closed in many facilities. Consequently, patient satisfaction has decreased, quality of care and patient safety have been compromised and the rate of medical errors has risen. Therefore, the nursing shortage has become one of the most pressing concerns for hospitals nationwide. For instance, the American Hospital Association estimates that 126,000 nursing positions are unfilled in the United States. The crisis will only deepen if the underlying causes are not addressed (Zangaro & Soken, 2007).

Turnover is defined as the number of resignation or termination divided by the average of direct and indirect care of registered nurses full-time equivalent (FTE) positions for the same year, which estimated to be 21.3% in 2010 (The HSM Group 2002). Nurse turnover has been linked to dissatisfaction (Hinshaw et al., 1987; Alexander et al., 1998; Tai et al., 1998; Larrabee et al., 2003), low salaries (Jones 1996), few years on the job and insufficient time to perform job tasks (Davidson et al., 1997), job possibilities elsewhere and supervisor behaviour (Tai et al., 1998), and some demographic characteristics such as being male and unmarried (Jones, 1996).

As nurses are the primary caregivers and are considered as the main professional part of the front line staff in most health systems, they play an essential role in delivering safe and effective care. At a time when a growing segment of the worldwide population ages and strains the capacity of nursing institutions, most are having difficulties in finding and retaining qualified nursing staff (Gohen & Van Nostrand, 1995). Saudi Arabia is no exception to the worldwide dilemma of the nursing shortage. However, the situation in Saudi Arabia is unique. In spite of nursing education having been available in Saudi Arabia since 1961, the numbers of Saudi nurses are only slowly increasing (El-Sanabary, 1993)

3 Proposed Methodology

This explorative descriptive study will examine the aspects of job stress, job satisfaction, burnout and intention to leave among nurses employed within the Ministry of Health hospitals in the Hail region of Saudi Arabia. The target population will be all available nurse managers and nursing staff who are working in all clinical areas (N = estimated 954). Prospective participants will be invited to participate by completing a survey. This project could assist to identify barriers of retention and seeks to make recommendations towards measures that will promote optimal retention rates and job satisfaction in Hail Region Hospitals. The first objective of the study is to examine the demographic and employment variables, job satisfaction, job stress, burnout among registered nurses. The second objective of the study is to investigate the nurses' intention to leave an organisation. The proposed method for this project consists of a survey of all available nursing staff working in the Ministry of Health hospitals in the Hail region of Saudi Arabia. This is a non-experimental explorative descriptive study. Such an approach is appropriate where new information is being sought from a population. A questionnaire and plain language statement will be distributed to the participants, via the nursing education department of each hospital. A total of 275 out of 954 nurses will be needed for the sample to be representative of the population. This is based on a confidence level of 95%, the confidence interval set at 5% (Creative Research Systems, 2007). Data was coded according to a numerical system transposed according to the nature of the questions. Frequencies, means and standard deviations will be analysed quantitatively using the most updated version of SPSS software. The SPSS software has proven to be consistently reliable in a variety of social science statistical analysis projects. It is currently the most widely used programs for statistical analysis in the social science. The Research Questions Guiding This Study Are:

- 1. What are the nurse's demographic characteristics and work profile characteristics, job satisfaction, job stress, burnout, and intention leave in a sample of nurses in selected hospital in Saudi Arabia?
- 2. What are the relationships between nurse job satisfaction, job stress and burnout?
- 3. Do the variables of job satisfaction, job stress and burnout contribute to nurses' intent to leave?

This research study is significant because it is the first empirical study of the factors that contribute to intention to leave in Saudi Arabia health services. Should this study identify these factors that influence nurse turnover rates, recommendations might be useful to decrease the rate of turnover and improve the retention of nurses. Patients, nurses, society and healthcare organisations will potentially benefit from the knowledge derived from the study. Moreover, retention of nurses is an important issue because shortages could lead to work overload, burnout and dissatisfaction of nurses, and compromised standards of patient care. Hospitals in Saudi Arabia

experience high turnover rates among professional nurses; therefore, reducing the turnover rates could save health care costs.

The nursing shortage is approaching crisis levels, and research in this area is important to all nurses and healthcare organisations. Furthermore, international nurses make sacrifices to work in Saudi Arabia and it is important to ensure that the workplace environment will be inviting and will not only reduce feelings of distinctiveness, but will meet the demands of the environment in which the nurses are practising. This study will focus on the nurses in the Ha'il region to understanding the factors that affect the nurses' job satisfaction, which is important in developing strategies for mitigating the nursing shortage.

4 Conclusions

For economical reasons, hospital administrators must be creative in finding methods to increase the nurses' job satisfaction, and by doing this; the financial stability of the hospitals will be significantly improved. However, there appears to be no studies that address the factors associated with nursing staff turnover issues in the Hail region's hospitals. Therefore, the purpose of this thesis is to explore the relationship between the work related variables of job stress, job satisfaction and burnout as predictors of nurses' intent to leave the job in the Hail region in Saudi Arabia as well as identifying the factors that affect nurses' turnover in their clinical care settings.

References

Alexander, J.A., Lichtenstein, R., Oh, H.J., & Ullman, E. (1998). A causal model of voluntary turnover among nursing personnel in long-term psychiatric settings. *Research in Nursing and Health*, 21(5), 415-427.

Abualrub, R.F. (2007). Nursing shortage in Jordan: what is the solution? J. Prof. Nurs., 23, 117-120.

Aiken, L.H, Clarke, S.P., Sloane, D.M., Sochalski, J., & Silber, J.H. (2002). Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA*, 288, 1987-1993.

Buchan, J. (2002). Global nursing shortages. BMJ, 324, 751-752.

Canadian Nursing Advisory Committee. (2002). Our Health. Our Future: Creating Quality Workplaces for Canadian Nurses. [Cited 4 Sep 2010.] Available from URL: http://www.hc-sc.gc.ca/hcs-sss/ pubs/nurs-infirm/2002-cnac-cccsi-final/index-eng.php.

Ceci, C., & McIntyre, M.A. (2001). A "quiet" crisis in healthcare: developing our capacity to hear. *Nurs. Philos.*, 2, 122-130.

Gohen, R.A., & Van Nostrand, J.E., (1995). Trends in the health of older Americans: United States, 1994. *Vital & Health Statistics*, 3(30), 1-339.

Hinshaw, A.S., Smeltzer, C.H., & Atwood, J.R. (1987). Innovative retention strategies for nursing staff. *Journal of Nursing Administration*, 17(6), 8-16.

Hirschfeld, M.J. (2009). Accepting responsibility for long-term care – a paradox in times of a global nursing shortage? *J. Nurs. Scholarsh.*, 41, 104-111.

Jones, C.B. (1996). Registered nurse turnover and the changing healthcare system. *Nursing Connections*, 9(1), 35-48.

Larrabee, J.H., Janney, M.A., Ostrow, L.C., Withrow, M.L., Hobbs, G.R., & Burant C. (2003). Predicting registered nurses job satisfaction and intent to leave. *Journal of Nursing Administration*, 33(5), 271-281.

Nikbakht Nasrabadi, A., & Emami, A. (2006). Perceptions of nursing practice in Iran. *Nurs. Outlook*, 54, 320-327.

Zangaro, G.A., & Soeken, K.L. (2007). A meta-analysis of studies of nurses' job satisfaction. *Res. Nurs. Health*, 30, 445-458.

Leadership styles of Nurse Managers at the Ministry of Health Hospitals of Saudi Arabia

Farhan Alshammari1

School of Health Sciences, RMIT University, Abuhala99@yahoo.com

Abstract. This explorative descriptive study will examine the leadership styles of nurse managers employed within the Ministry of Health hospitals in Saudi Arabia. It also will investigate nursing staff levels of job satisfaction, effectiveness and willingness to exert extra effort at their workplace. There will be a two phase non experimental explorative descriptive design employing quantitative methodology. Both Managers and staff will be surveyed using the Multifactor Leadership Questionnaire (MLQ). Local and expatriate nurse managers will be invited to participate in the study. General nursing staff of wards whose manager elected to participate will then be invited to be surveyed. The total population of nurses in the participating hospitals is estimated at 954. This study builds on past research studies and adds to the growing body of knowledge on the transformational leadership theory and leadership outcomes which has been introduced by Burn (1978) and developed by Bass (1985). It adds to the body of knowledge in leadership research in the nursing profession and also contributes to better understanding of the way leadership is enacted in a sub-populate area like Saudi Arabia where nurses are from various and different cultures. This study will provide the nurse managers with the opportunity to examine their own leadership style and reflect on how they affect their own nursing staff.

Keywords: Leadership Styles, Nursing leadership, Job satisfaction, Transformational leadership, Nurse Manager)

1. Introduction

Leadership plays a very crucial role in the human affairs, and it has been a favorite subject for researchers. Marquis and Huston (2010) claimed that the work environment and the work performed by people in any organization are highly affected by the leadership style of its managers. Any profession requires a leadership, as it is the case with all the health care sectors. Nurses are the key role in shaping the nursing professions; nurses who work at the heath division must have the required

knowledge to provide appropriate heath care and accept the responsibilities of professional management and leadership. They need to perform a supportive and coordinative role; ensuring the continuity of patient care and thereby making a difference in the process of healthcare service provision (Duygulu & Kublay, 2011).

2. Literature review

Leadership theories have evolved as time has progressed. Various scholars have described the evolution of leadership theories: Bolden et al (2003) and Casida (2007). According to Casida (2007), the evolution of leadership theories started with the great man theory (pre-1900s), and then followed by the trait theory between 1900 and 1948. This was subsequently followed by the contingency theory between 1948 and 1980. This was consequently followed by the transformational leadership theory which is presently being utilized. Casida also identifies other theories that are being utilized presently as servant and multifaceted leadership theory.

There are many different types of leadership that nurse managers and leaders demonstrate to lead staff nurses in hospitals such as (autocratic, democratic, laissez-faire, bureaucratic and situational) and contemporary leadership (charismatic, transactional, transformational, connective and shared leadership) (Huber, 2006). It is believed that in hospital settings leadership style of head nurses might affect the job satisfaction (AL-Hussami, 2008). Bass and Avolio (1990) found that transformational (TF) leadership styles were preferred over transactional (TA) leadership styles, and managers who exhibited transformational characteristics reported more satisfied staff nurses.

According to Albejaidi (2010), the Saudi Arabian government has taken the lead to improve the health care system and has been funding it over the years. As a result of this, Albejaidi notes noted that the World Health Organization's (WHO) ranked Saudi Arabia healthcare system the 26th in the world, outweighing other major economic countries like Canada, and the United States of America. Saudi Arabia is developing very fast in all disciplines, especially in nursing and health (Aldossary et al. 2008).

Nevertheless, the demand for labor in the health sector cannot be met by Saudi nationals alone because of the relatively small number of Saudi graduates from nursing schools (Abu-Zinadah, 2004). Consequently, most of the healthcare providers are expatriate from different countries. For example, nurses come from over forty different countries, including the United Kingdom, Ireland and the USA (Aldossary et al. 2008). The nursing leaders in Saudi Arabia faced several problems related to this mixed workforce such as recruitment and retention problems, high turnover rate, and cultural difference. Successful, motivated, and visionary leaders will be required in order to effectively manage Saudi Arabia's diverse nursing workforce. The current and future rapid developing health care environment in Saudi Arabia demands skilled and competent leaders who are able to inspire their staff with a vision of what can be accomplished no matter the challenges.

3. Proposed Methodology

There will be a two phase non experimental explorative descriptive design employing quantitative methodology. The study aims to collect data from two groups: Phase one: Ward/Unit Nurse Manager and Phase two: Nursing Staff. The participants are working in five of the Ministry of Health Hospitals in Saudi Arabia. A total population of N= (954) nurses. The research instrument is Multifactor Leadership Questionnaire 5X Short (MLQ). Leader and Rater form will be used in this Study.

The research questions guiding this study are:

- What leadership styles are practiced by nurse managers in the Saudi Ministry of Health hospitals?
- What are the relationship between nurse mangers leadership styles and their effectiveness?
- Is there an association between the perceived leadership style of the nurse manager and the subordinates' willingness to exert extra effort and experience job satisfaction?
- Is there a difference between the nurse managers' perceptions of their own leadership style with how the manager's subordinates perceive the manager's leadership style?

4. Data Collection Approach

Numerical Codes will be utilised to identify each of the 5 hospitals participating in the survey ranging from (1 to 5). Each ward/unit managers will be allocated an alphabetical code for example hospital 1 ICU unit (1B). Once the Nurse Manager elects to participate by returning their survey, their staff will be invited to participate using the hospital /ward code. These codes will be written in the space allocated for the 'Name of Leader' in the first section of the MLQ form. This will help to match the nurse manger survey with his/her staff surveys. It is anticipated that a total sample of 954 Nurse Managers and nursing staff will be invited to participate in the study.

5. Data Analysis

SPSS version 19 will be used for data analysis. Descriptive and inferential statistic will be provided.

6. Conclusion

This paper draws attention to the nursing leadership styles practiced by nurse manager in the ministry of health hospitals in Saudi Arabia and how the transformational leadership is very essential for the achievements of the health care sectors. The expected study result will reflect the dominant leadership style practiced by the nurse manager and how it's affect the staff satisfaction and willingness to exert extra effort.

References

Abu-Zinadah, S (2004) The situation of Saudi nursing. Health Forum, 52, 42-43.

Albejaidi, M. F (2010). *Healthcare system in Saudi Arabia: An analysis of structure, total quality management and future challenges.* Retrieved August 18, 2011, from http://www.qualityandsafety.info/article.pdf

Aldossary, A., While, A. & Barriball, L (2008) Health care and nursing in Saudi Arabia. *International Nursing Review*, 55, 125–128.

AL-Hussami, M (2008) Study of nurses job satisfaction: The relationship to organizational commitment, perceived organizational support, transactional leadership, transformational leadership, and level of education. *European Journal of Scientific Research.* 22, 286–295.

Bass, B.M (1985) *Leadership and performance beyond expectations*, Free Press, New York.

Bass B. & Avolio B (1990) The implications of transactional and transformational leadership for individual, team, and organizational development. *Research in Organizational Change & Development* 4, 231–272.

Bolden, R., Gosling, J., Marturano, A., & Dennison, P (2003) A review of leadership theory and competency frameworks, University of Exeter, Dunsford Hill.

Burns, J. M (1978). Leadership, Harper & Row, New York.

Casida, J. M (2007) *The relationship of nurse managers' leadership styles and nursing unit organizational culture in acute care hospitals in New Jersey*, Seton Hall University, New Jersey.

Duygulu, S. & Kublay, G (2011) Transformational leadership training programme for charge nurses. *Journal of Advanced Nursing*. 67, 633–642

Huber, D (2006) *Leadership and Nursing Care Management*, 3rd edn, Saunders Elsevier, Philadelphia, PA.

Marquis, BL. & Huston, CJ (2010) *Leadership Roles and Management Functions in Nursing: Theory and Application*, Lippincott Williams & Wilkins, Philadelphia, PA.

Nurses' knowledge regarding pain management in Hail Region Hospitals, Saudi Arabia

Hamdan Albaqawi

Bcs, RN, MSN, PhD candidate (School of Health Sciences), RMIT University, Wesaam_2003@yahoo.com

Abstract. Pain is a human experience that affects overall quality of life and one of the most common reasons for people seeking health care. Effective pain management requires precise knowledge, attitude and competent assessment skills. This explorative descriptive mixed methods study will be conducted in Saudi Arabia (SA) and will include sample of local and expatriate nurses who are working in Hail Region Hospitals. The study seeks to identify the nurses' knowledge and attitudes regarding pain management using the "Knowledge and Attitudes Survey Regarding Pain" (NKASRP). Possible barriers to achieving optimal pain management will be explored. The study consists of two phases and will include survey and semi-structured interview. Quantitative data will be analysed by using SPSS and qualitative data will be analysed by using thematic analysis.

Keywords: Pain assessment, Pain management, Nurses' Knowledge, Nurses' attitudes, Saudi Arabia.

1 Introduction

Pain is a part of life that everyone must experience at some point during their lives. It is the main reason people seek health care (Polomano, Dunwoody, Krenzischek & Rathmell, 2008), and is one of the most common reasons that patients seek health care in a hospital setting (McLean et al., 2004). An estimated 20%-40% of the patients treated by general practitioners are found to suffer from different pain conditions (Hasselstrom, Liu-Palmgren & Rasjo-Wraak, 2002). In Australia, it is estimated that one in five people (about 3.2 Million Australians) including children and adolescents, will suffer chronic pain in their lifetime (Walsh et al., 2008). Pain assessment and management is the most fundamental part of the nurse's responsibility when it comes to a patient with pain (Innis, Bikaunieks, Petryshen, Zellermeyer & Ciccarelli, 2004). However, this could be highly influenced by the nurses' knowledge and attitude regarding pain management in Hail Region Hospitals in Saudi Arabia as well as seek to identify possible barriers to effective pain management.

2 Literature Review

The primary responsibility of health care professionals, especially nurses, is to relieve the pain and suffering of patients. It is the moral and ethical responsibility of the nurse and a fundamental human right for patients to live free of pain (Cousins et al., 2004). Although pain can be effectively treated and relieved, undertreatment of pain remains a significant

clinical problem and it has been recognised as an area of concern among health professionals, patients, and health care organisations. Even in hospital settings where pain should be treated effectively, research evidence shows that pain is managed inadequately and there are a large number of patients who still suffer from unrelieved pain (Dolin, Cashman, & Bland, 2002) very old reference. Despite the development of new techniques in managing pain, many patients continue to suffer from pain (Solman, et al, 2005; Solman, et al., 2006; Pasero & McCaffery, 2007). Pain has often been poorly assessed and inadequately managed and the undertreatment of pain has been reported for many decades as a major and persistent clinical problem (Brown, Bowman, & Eason, 1999; Fosnocht, Swanson, & Barton, 2005; McCaffery & Ferrell, 1997; McCaffery & Pasero, 1999; Schafheutle, Cantrill, & Noyce, 2001). Jones et al. (2004) identified that nurses have knowledge deficits and incorrect beliefs about pain assessment and management that affect the way the patients' pain is managed. Thus, these misconceptions and deficits can leads to inappropriate, incorrect, and inadequate pain management practices (Mezey, 2005; Twycross, 2002). Lack of knowledge about pain and pain treatment, and myths about addiction is considered a significant barrier to effective pain management. Numerous studies indicate that nurses are not managing pain properly due to deficits in their knowledge and beliefs (Jones et al., 2004; Mezey, 2005; Twycross, 2002). These studies have identified notable knowledge deficits and incorrect beliefs among nurses about pain assessment and its treatment. Furthermore, lack of knowledge about pain and its treatment have been stated as significant obstacles to effective pain management (Jones et al., 2004). A number of researchers have indicated that nurses are worried about the possibility of addiction and consequently they underestimates patients' pain (Schafheutle, Cantrill & Noyce, 2001). Many studies reveal that some nurses are reluctant to administer opioids based on negative attitudes (Drayer, Henderson, & Reidenberg, 1999; Edwards et al, 2001). Other studies found that some nurses aimed to reduce pain rather (Edwards et al., 2001; Twycross, 2002). In spite of numerous studies identifying the deficit in general pain management knowledge the problem remains that patients continue to suffer from unnecessary pain (Solman, et al., 2005; Solman, et al., 2006; Pasero & McCaffery, 2007).

The consequences of pain mismanagement result in both human suffering and economic costs (Brennan et al., 2007; Innis et al., 2004; Maclaren & Cohen 2005). Pain is the third most costly health problem in Australia and it costs the Australian economy over \$34.3 billion per annum, or \$10,847 per person affected (Cousins, Bridenbaugh, Carr, & Horlocker, 2008). In the United States, Chronic pain is affecting more than the third of the American population and the annual cost of it estimated to be \$100 billion (National Institutes of Health, 1998). On the other hand, adequate pain assessment and management may reduce medication cost, improve patient outcomes and satisfaction with care and shorten hospital stays (Innis et al., 2004; Polomano et al., 2008).

3 Method

This study consists of two phases; the first phase will utilise a survey questionnaire "NKASRP" that will be administered to nursing staff working in Hail Region Hospitals. The questionnaire will be distributed to the participants (N=275) to determine their knowledge and attitude regarding pain. The "Knowledge and Attitudes Survey Regarding Pain" tool can be used to assess nurses and other professionals in your setting and as a pre

and post test evaluation measure for educational programs. The tool was developed in 1987 and has been used extensively from 1987 - present.

The second phase will utilise a qualitative methodology (semi-structured interviews (N= 20-30) to explore the barriers to achieve optimal pain management as perceived by nurses. All interviews in this phase will be audiotaped. The inclusion criteria for the selection of participants will be: nurses aged from 21-65 years old, who have more than 6 months working experience and are able to read, write and comprehend English.

The research questions are:

- What knowledge and attitudes do nurses hold regarding pain management in Hail Region Hospitals?
- What are the barriers to achieving optimal pain management as perceived by nurses working in Hail Region Hospitals?
- What demographic and cultural factors impact on the delivery of effective pain management?

3.1 Instrument

The "Knowledge and Attitudes Survey Regarding Pain" (NKASRP) questionnaire is developed by Betty Ferrell and Margo McCaffery. The questionnaire consists of 22 true or false items, 13 multiple choice items, and 2 case studies with 2 multiple choice items each. This tool has been used widely and designed to measure knowledge and attitudes of health care professionals regarding pain. The tool was developed in 1987 and has been used extensively since 1987. The content of the tool was extracted from current standards of pain management such as the World Health Organization, the American Pain Society, and the Agency for Health Care Policy and Research.

3.2 setting

Saudi Arabia (SA) is a developing country located in the Middle East and occupies four fifths of the Arabian Peninsula. The Kingdom's population is estimated to be around 25 million including 21.6% non Saudi born citizens, whilst 90% of the Saudi citizens are Arabs and all are Muslims (Central Intelligence Agency, 2010). The total area of the Kingdom is 2,240,000 square kilometres (Al-Shahri, 2002). The capital city of KSA is Riyadh, and the country is divided into thirteen different provinces or regions .The study will be conducted in five Hospitals in Hail Region which located in the north of Saudi Arabia. There are 13 hospitals in the region, 4 central hospitals and 9 peripheral hospitals including general, maternity and psychiatric hospitals. The surveys will be distributed in 5 general hospitals two are cental and three are perepheral. Hospital A is the largest hospital in the region with capacity of 210 beds. Hospital B is the second largest hospital with capacity of 200 beds. Hospital with capacity of 50 beds. Hospital E is a peripheral hospital with capacity of 30 beds.

3.3 Justification of the study

Although many studies investigated the nurses' knowledge and attitudes towards pain management; these studies were conducted generally in western countries. Few studies have investigated knowledge and attitudes towards pain in the Middle East and namely in Saudi Arabia. This study is significant as no other studies have been conducted in Hail Region Hospitals to determine the nurses' knowledge and attitudes regarding pain management. This is most interesting since there are many expatriate nurses working in these hospitals and has left a large gap in the area of research investigating these nurses' knowledge and attitudes. The value of this study is that it provides information about nurses' level of knowledge regarding pain management and attitudes toward pain. This study can help to identify the barriers to effective pain management by these nurses.

4 Conclusions

Nurses have the primary responsibility and accountability to assess the patient's pain, implement intervention, and evaluate the interventions. However, undertreatment of pain has been reported in many studies as a persistent clinical problem. Effective pain management could result in reducing the human suffering and medications cost. This study will explore pain management practices in one region of Saudi Arabia and has the potential to improve the nursing health care of the patients in these hospitals and will add to the body of knowledge regarding pain management.

References

Al-Shahri, M. (2002). Culturally sensitive caring for Saudi patients. *Journal of Transcultural Nursing*, 13(2), 133-138.

Brennan, F. Carr, D. Cousins, M. (2007) Pain management: A fundamental human right. *Anesthesia and Analgesia*, 105, 205-221.

Brown, S. T., Bowman, J. M., & Eason, F. R. (1999). Assessment of nurses' attitudes andknowledge regarding pain management. *Journal of Continuing Education in Nursing*, *30*, 132-139.

Central Intelligence Agency. (2010, November 9). *The World Factbook: Saudi Arabia*.[website](https://www.cia.gov/library/publications/the-world-factbook/geos/sa.html, accessed 11 November 2011).

Cousins, M. J., Bridenbaugh, P. O., Carr, D. B., & Horlocker, T. T. (2008). *Cousins and Bridenbaugh's Neural Blockade in Clinical Anesthesia and Pain Medicine*. Philadelphia, Lippincott Williams & Wilkins.

Cousins, MJ., Brennan, F. & Carr, DB. (2004). Pain relief: a universal human right. *Pain*, 112, 1-4.

Dolin, S. J., Cashman, J. N., & Bland, J.M. (2002). Effectiveness of acute postoperative pain management: Evidence from published data. *British Journal of Anaesthesia*, *89*, 409-23.

Edwards, H. E., Nash, R. E., Najman J. M., Yates, P. M., Fentiman, B. J., Dewar. A., et al. (2001). Determinants of nurses' intention to administer opioids for pain relief. *Nursing and Health Sciences*, 3(3), 149-159.

Fosnocht, D. E., Swanson, E. R., & Barton, E. D. (2005). Changing attitudes about pain and pain control in emergency medicine. *Emergency Medicine Clinics of North America*, 23, 297-306.

Hasselstrom, J., Liu-Palmgren, J., Rasjo-Wraak, G. (2002). Prevalence of pain in general practice. *Eur J Pain, 6,* 375–85.

Innis, J., Bikaunieks, N., Petryshen, P., Zellermeyer, V., & Ciccarelli, L. (2004). Patient's satisfaction and pain management: An educational approach. *Journal of Nursing Care Quality*, 19(4), 322-327.

Jones, K. R., Fink, R., Pepper, G., Hutt, E., Vojir, C. P., Scott, J., Clark, L., & Mellis, K. (2004). Improving nursing home staff knowledge and attitudes about pain. *The Gerontologist*, 44(4), 469-478.

MacLaren, J. Cohen, L. (2005) Teaching behavioural pain management to healthcare professionals: A systematic review of research in training programmes. *The Journal of Pain*, 6(8), 481-492.

McCaffery, M., Ferrell, B. R. (1997). Influence of professional vs. personal role on pain assessment and use of opioids. *Journal of Continuing Education in Nursing*, 28(2), 69-77.

McLean, S., Domeier, R., DeVore, H., Hill, E., Maio, R. & Frederiksen, S. (2004). The epidemiology of pain in the pre-hospital setting. *Prehospital Emergency Care*, 6(4), 402-405.

Mezey, M. (2005). Assessing pain as the fifth vital sign in ling-term care facilities: Recommendations from the field. *Journal of Nursing*, *31*(3), 16-24.

National Institutes of Health, (1998). *NIH Guide: New Directions in Pain Research*: I. Bethesda, MD: National Institutes of Health. Retrieved from: http://grants.nih.gov/grants/guide/pa-files/PA-98-102.html

Pasero, C, & McCaffery, M. (2007). IV opioid range orders for acute pain management. *American Journal of Nursing*, 107(12), 52-60.

Polomano, R. Dunwoody, C. Krenzischek, D. Rathmell, J. (2008) Perspective on Pain Management in the 21stCentury. *Pain Management Nursing*, 9(1), 3-10.

Schafheutle, E. I., Cantrill, J. A., & Noyce, P. R. (2001). Why is pain management suboptimal on surgical wards? *Journal of Advanced Nursing*, *33*, 728-737.

Solman, R., Rosen, G., Rom, M., & Shir, Y. (2005). Nurses' assessment of pain in surgical patients. *Journal of Advanced Nursing*, 52(2), 125-132.

Solman, R., Wruble, A., Rosen, G., & Rom, M. (2006). Determination of clinically meaningful level of pain reduction in patients experiencing acute postoperative pain. *Pain Management Nursing*, 7(4), 153-158.

Twycross, A. (2002). Education nurses about pain management: the way forward. *Journal of Clinical Nursing*, *11*, 705–714.

Walsh, N. E., Brooks, P., Meike Hazes, J. M., Walsh, R. M., Dreinhöfer, K., Woolf, A. D., Ăkesson, K., & Lidgren, L. (2008). Standards of Care for Acute and Chronic Musculoskeletal Pain: The Bone and Joint Decade. *Archives of Physical Medicine Rehabilitation*, 89, 1830-45.

Deception and its possible relation to emotional intelligence among respondents in the Czech Republic

Lenka Mynaříková1

¹ Department of Psychology, Charles University in Prague, l.mynarikova@centrum.cz

Abstract. In the presented study we oriented on finding a possible relation between deceptive behavior and emotional intelligence Czech Republic. Emotional respondents in the among intelligence is related to self-deception and deception detection, vet there is not enough information on its possible relation to lying behavior itself. We used a deception behavior survey to analyse deception characteristics in the Czech Republic and the TEIQue inventory to measure emotional intelligence. ANOVA and regression analysis revealed significant differences among respondents in lying behavior, which might be influenced by the global trait of emotional intelligence or particular factors as well-being. The estimated number of lies negatively correlated with the emotional intelligence scores, suggesting that respondents with lower emotional intelligence say less lies and also tend to lie to adults more often than to children. These results might be beneficial for interventions in pedagogical or forensic psychology or in family therapy and can improve our understanding of variables affecting lying behavior.

Keywords: *deception, emotional management, well being, interpersonal relationships*

1 Introduction

It was the American sociologist Goffman who – describing the theory of social interactions - said that we all are just players in a theater game. We put various masks on our faces in order to play a diversity of characters and in order to cast a desired impression on others. By playing with illusions we might say that we are somehow deceiving others, because we are not showing our real self, but a variety of selves, which corresponds the best with the current situation. Deceit has become a part of our daily existence and it has become a regular way of communication between each

other, even though people only unwillingly admit they lie. Deception, its role in society and in the interpersonal communication and relationships and the question of its morality are subjects of interest among psychologists, philosophers, economists or even fiction writers.

The literary view on deception described Kafka (2003) in his novel Trial. He said lying was a universal principle and lies and concealment were the foundation stones of most contemporary societies. Similary Beckett (1980) and Eco (1999) thought of lies as a type of a contemporary agenda, a general semiotic structure. Defoe described fictional literature in general as a type of deception and writers as the masters of lying (Smyth, 2002).

Moral aspects of deception have been a subject of an intensive debate among Western philosophers for many centuries. From the philosophical point of view deception is often a matter of an immoral character and the situational and dynamical aspects are completely omitted (Nyberg, 1993; Lewis, & Saarni, 1993). This extremist opinion was taught for example by Kant (1949), who said that we had to tell the potential murderer where to find his victim, because the crime itself would not outweigh the immorality of lying. On the contrary, Bentham or Mill believed in the utilitarian perspective, which emphasized the importance of context information on the evaluation of the moral consequences of lying (Fu et al., 2007). Sometimes deception can be desirable and supported according to those philosophers.

The studies on deception have a long tradition also among economists. The economical theory of lying is based on the concept of 'homo economicus', i.e. a selfish person uninterested in other people good (Becker, 1993). Such a person lies, whenever it is beneficial for them, regardless of the impact of their lies on others (Gneezy, 2005). Founders of the theory of games also contributed to the explanation of lying, when they studied bluffing among poker players (Ferguson, & Ferguson, 2003).

Finally the psychological point of view analyzes deception from many aspects like the frequency of lying, age and gender differences among liars, relationship between the liar and the deceived person, or types of lies and their functions. Lying as a psychological concept takes into account interpersonal relationships and social interactions in general and it does not anticipate any moral conclusions. Lying is described as a social activity which serves as a connection among people and which cannot be simply described as wrong or right. It stresses the importance of the situation in which deception occurs, the effect of other involved persons and of liar's intentions.

Whereas the topic of the characteristics of deceptive behavior has already attracted a lot of interest in the foreign countries, experts in the Czech Republic generally pay only marginal attention to deception, mostly in relation to moral development and communicative and manipulative strategies. The research question is whether deceptive behavior in the Czech Republic manifests in the same way as was described in foreign countries, especially in the USA. Moreover we tried to find out whether there was a relationship between deception and emotional intelligence as a relationship between emotional intelligence and deception have already been found in a couple of studies (Mynaříková, & Boukalová, 2012).

2 Literature Review

2.1 The frequency of lying

The documented frequency of lying ranges between one or two lies a day (DePaulo et al., 1996; Hancock, Thom-Santelli, & Ritchie, 2004) and two lies told during each 10 minutes long conversation (Tyler, Feldman, & Reichert, 2006). The respondents in Hancock's research (2007) admitted to lie in 14% of e-mails, 27% of face-to-face interactions and in 37% of phone calls. Other studies show these numbers are even higher. In 1995, 90 % of participants of an extensive research (Patterson, & Kim, 1995) admitted they lied. Boon and McLeod (2001) add that under certain circumstances people consider lying justifiable and even though we are longing for honesty in our relationships, we lie about our income, successes, education, age or sexual relationships almost on a daily basis (Ford, 1999).

With such a high frequency it is interesting that similar researches are missing in either our country or in many other middle European countries. The question is whether the situation in the Czech Republic is similar to the situation in the USA and whether the possible differences exist also among particular countries in the middle Europe.

2.2 Lying in relationships

According to Lewis and Saarni (1993) we lie more often to our intimate partners and family relatives than to foreign people. On the contrary DePaulo and Kashy (1998) consider lies to foreign people the most frequent. The frequency of lying to foreigners and to close persons (best friends and intimate partners) might depend on the type of attachment we tend to maintain in social relationships (Ennis, Vrij, & Chance, 2008). People with insecure attachment are thought to lie more often to their best friends and to foreigners. People with avoidant attachment may lie more often to their intimate partners.

With the development of a relationship the ability of partners to lie to each other develops as well (Cole, Leets, & Bradac, 2002; Peterson, 1996). Partners have a chance to find out the best way of talking and behaving in order to deceive each other. The so-called ostrich effect also helps with deception - in an attempt to maintain the relationship and to avoid endangering factors, people sometimes tend to ignore even the most obvious lies (Vrij, 2008). In certain situations it may be more pleasant to hear a lie than a truthful fact.

2.3 Gender aspects of lying

Men lie as often as women, but they say different types of lies (DePaulo et al., 1996). Women say lies oriented to others, especially to other women, and they lie mostly about their feelings or opinions in order not to hurt others' feelings and in order to avoid excessive criticism (DePaulo et al., 1996; DePaulo, & Bell, 1996). Men say more self-oriented lies to present themselves as more capable or educated (DePaulo et al., 1996).

Gender differences in lying on the internet were also examined. Between men and women there are no significant differences in the frequency or in the reasons behind lying (Caspi, & Gorsky, 2006; Whitty, & Gavin, 2001). Whitty (2002) drew attention to the relation between the age of men and women and the frequency of lying – men older than 30 years lied as often as adolescent men, women lied less often after they had reached 30 years of age.

Gender differences in deception can be partially explained by the different parental approach to upbringing girls and boys (Feldman, Tomasian, & Coats, 1999; Westcott, Davies, & Clifford, 1991). Girls are more praised for showing their emotions and they have more opportunities to train nonverbal expression of emotions. As they grow older girls learn to reveal less signs of lying in their faces and become better liars (Vrij, 2008). Boys, who are led to hide negative emotions, are trained in neutralizing them. With increasing age boys are more capable of pretending neutral feelings about something they do not like (Vrij, 2008).

The differences in deception between men and women are influenced by hormones (Croson, & Gneezy, 2004). Oxytocin increases the level of trust (Kosfeld et al., 2005), estrogen decreases the willingness to risk (Chen, Katuscak, & Ozdenoren, 2005) and testosterone influences the tactics of behavior (Apicella et al., 2008; Van den Bergh, & Dewitte, 2006).

2.4 Developmental aspect of lying

As we have seen, some of the differences in lying between men and women may be described with the help of developmental and educational theories. It may be surprising that even very small children are able to deceive others and they use lies in their everyday communication with almost the same frequency and same success as their adult opposites.

Lately this particular aspect of deception theories has gained a lot of significance especially in relation to moral development of children, the theory of mind and the attachement theory (Talwar, & Lee, 2002).

In order to lie, children must have a developed self-referential system, that helps them realize what they want to do and how it would influence other persons. The ability to lie appears probably no sooner than at the age of two (Lewis, & Saarni, 1993). Children at the age of two and a half are able to deceive others by concealing an information, especially when asked to lie by an adult (Newton, Reddy, & Bull, 2000; Reddy, 2007; Wilson, Smith, & Ross, 2003). More difficult falsifications appear later (Ekman, 1992). Using diary recordings, Newton, Reddy and Bull (2000) found out that three-year-olds lied and used various forms of deception including excuses and accusations. In a longitudinal study of 40 Canadian families (Wilson, Smith, & Ross, 2003), there were 35 % of two-year-olds, 15 % of four-year-olds and 5 % of six-year-olds, who did not lie at all. The lies of pre-school children are mostly given away by their verbal behavior. Preserving the consistency between the primary lie and its

repetitions is more difficult for children younger than seven years of age (Talwar, Gordon, & Lee, 2007).

Children learn to lie by imitating their parents' behavior (Lewis, & Saarni, 1993). To save other people's feelings and retain good social relationships, children are led to say prosocial lies in complimentary situations (Talwar, Murphy, & Lee, 2007). Whilst three-year-olds are not experienced in this and tell the truth even in an inappropriate situation, older children are good at using polite lies in order not to insult someone or disrupt good mutual relationships (Talwar, Murphy, & Lee, 2007). The adequate use of prosocial lies appears between the forth (Bussey, 1999) and the fifth year of age (Talwar, Murphy, & Lee, 2007).

2.5 Intercultural aspects of lying

Lying is a widespread phenomenon among men and women of all ages. Researches have proved that this applies worldwide and that although there are differences in lying among various cultures, we can meet with lies everywhere.

The cultural differences in lying are related to the universality and cultural specificity of moral concepts and judgements. The available research has not provided enough of clearly interpreted results yet and only few of the existing researches paid attention to other than Eastern society (Fu et al., 2001; Fu et al., 2007; Lee et al., 1997). Fu et al. (2007) compared the concepts of lying among Chinese and Canadian children. Both groups were successful in recognizing truthful and deceitful statements, but Chinese children preferred saying a lie to help the group even if it hurts the individual, whilst Canadian children preferred the other situation when helping an individual had a negative influence on the group (Fu et al., 2007). The significant differences in estimating what is and what is not a lie are found also in adult population – Chinese students do not consider it a lie, when they conceal their own good deeds, whilst Canadian students consider it lying (Fu et al., 2001).

The different concepts of lying in Canada and China reflect two dimensions of societies. In collectivistic societies as American Samoa, people are more willing to lie if their lie is related to group or family matters (Aune, & Waters, 1994). However, the inhabitants of the more individualistic North America lie more often to protect their own privacy or feelings of close relatives (Aune, & Waters, 1994). Sweet et al. (2010) confirmed similar results when they compared children from collectivistic China and individualistic USA.

The importance of lying is more significant in some cultures than in the others. Mealy, Stephan and Urrutia (2007) found out that for people in Ecuador lying was significantly more negative than for the inhabitants of the USA.

2.6 Typological and functional aspects of lying

Deception exists in various forms, as we can notice every day. Some lies are more serious than others and their consequences also greatly differ. We often use various types of lies due to the function which the lie should serve for.

One of the possible typologies is dividing lies by their relevance (Vrij, 2008). Social lies are told on a daily basis in order to present ourselves or our close ones in a better light, not to hurt others or to pleasure others. Serious lies belong to the forensic field, because they occur in the interrogation of crime suspects/victims/witnesses. This typology is not completely clear, because some lies do not belong to the category of social lies, but they still cannot be considered serious lies (e.g. lying about infidelity in intimate relationships).

The second most often used typology divides lies by the amount of modifications made in the previously truthful information (Granhag, & Vrij, 2005). Falsifications are complete lies that are an absolute opposite of truth, distortions are small deflections from truth, and concealments appear, when we say we do not remember or know something. Metts (1989) divides lies into contradictions, distortions and omissions, but these terms correspond with those previously mentioned. DePaulo et al. (1996) use the terms direct lies, exaggerations (over- or underestimations) and subtle lies. Bussey (1999) divides lies into antisocial lies told in order to deny guilt and to avoid problems, social or 'white' lies told in order not to hurt other people and tricks told for entertainment. Bok (1982) describes professional, romantic, political, therapeutical, medical or experimental lies.

Lying to others is usually typically based on motives as avoiding punishment, acquiring a reward, protecting someone against punishment, gaining admiration from others, getting out of an unpleasant situation, avoiding embarrassment, keeping a secret or winning control over others (Gudjonsson, 1994). The most common reasons for children lies are similar, but they especially lie to avoid the responsibility for an offense, falsely accuse their siblings and gain control over others' behavior (Wilson, Smith, & Ross, 2003). The 'duping delight' concept explains why some people lie even when it is not beneficial for them – they experience a pleasure in deceiving others and in not being caught (Vrij, 2007).

DePaulo et al. (1996) studied diary recordings of university students and found out that motives for lying could be categorized into three dimensions: lying for the good of oneself or others, lying for gaining benefits or avoiding costs and lying from material or psychological reasons. Mancas (2007) later added the fourth, esthetic dimension, in which a person uses lies to create alternate worlds.

Respondents in researches often do not find social lies to be lies at all, therefore we can expect that prosocial lies would be more often considered innocent than serious in our research. It is also highly probable that the reasons of men's and women's lies would not differ.

2.7 Deception and emotional intelligence

Emotional intelligence involves intrapersonal and interpersonal factors. Similarly, deception is intra- and interpersonal phenomenon, as we have seen above. There are two general models of emotional intelligence. The first one thinks of emotional intelligence as of a mental ability, related to the management of emotionally relevant information. It includes 4 components – perception of emotions, emotional support of thinking, emotional understanding and emotional management (Mayer, & Salovey,

1997). In this concept emotional intelligence is tested with maximum performance tests. The second model tries to find concrete abilitites, behavioral tendencies and personality traits. It stresses the importance of those characteristics, that are responsible for individual differences in processing of the emotional information (Petrides, 2009). Emotional intelligence in this model is tested by self-report inventories.

The relationship between deception detection has already been proved (Costanzo, & Archer, 1993; Mynaříková, & Boukalová, 2012; Porter, ten Brinke, Baker, & Wallace, 2011). A number of studies also documented the relationship between self-deception and emotional intelligence (Erez, Johnson, & Judge, 1995; Mannion, 2009). Dunn and Hughes (1998) stressed the developmental aspect of the relationship. They found out that four-year-old children, who scored higher in deception and emotional understanding (one factor of emotional intelligence), were better at 'reading' and differencing emotions 7 months later. Understanding the role of self-deception in emotional intelligence of emotional intelligence in the deception detection is an interesting issue for forensic psychologists. The relationship between emotional intelligence and deception of others has not been studied so far, but since deception is connected with the lie detection and it has common basis with self-deception, we might expect there those two concepts would be also in a relation.

3 Data and Methodology

3.1 Respondents

There were 112 respondents (79 women, 33 men), who participated in the research (see Table 1). All of them were older than 15 years, born and living in the Czech Republic. There were 5 age groups: 15 - 25 years, 26 - 35 years, 36 - 45 years, 46 - 55 years, 56 and more years. Among the respondents there was a variety of students of pedagogy, science, mathematics, primary school teachers, administrative workers or doctors.

		Frequency	Percent	
Age	15-25	57	50,9	
	26-35	29	25,9	
	36-45	10	8,9	
	46-55	12	10,7	
	56 and more	4	3,6	

Table 1. Demographic characteristics of the sample

		Frequency	Percent
Age	15-25	57	50,9
	26-35	29	25,9
	36-45	10	8,9
	46-55	12	10,7
	56 and more	4	3,6
	Total	112	100,0

3.2 Materials

For collecting the data we asked the participants to fill in an online survey designed for a previous research (Mynaříková, 2012). Each item was chosen after an in-depth analysis of foreign literature and research studies and covered various aspects influencing deception: a developmental aspect (age), gender, functional aspects, the frequency of lying, and the role of relationships in lying. Questions with one or multiple possible responses and open-end questions were used.

In the second phase respondents filled in the Trait Emotional Intelligence Questionnaire (TEIQue), which is available, free of charge, for academic and clinical research.

4 Empirical Analyses

All respondents admitted to lie at least sometimes, with 90 % of them telling 1 - 12 lies a week (a self-estimation). Distribution of the estimated number of lies told in a week ranged between 1 - 122 (M= 10.79, SD = 22.25, Min = 6, Max = 122 for men; M = 4.62, SD = 6.99, Min = 1, Max = 52 for women). The difference between men and women is significant, t(112) = 2.226, p < .05.

Yet for 100 % of participants honesty – contrary to the sense of humor, tolerance or similar hobbies - is highly important for maintaining relationships. This result was independent on gender or age. For 56.3 % of respondents lying is rarely or never justifiable, independent on gender of respondents, F(110,1) = .671, p > .05, but dependent on their age, F(107,4) = 3.484, p < .01, with the younger respondents justifying lies more often than the older respondents.

The estimated number of lies told in a week correlated significantly with self-control (r = -.180, p < .05) and global trait of emotional intelligence (r = -.158, p < .05). Both correlations remained significant when controlled for gender (r = -.184, p < .05, and r = -.195, p < .05 respectively). Other TEIQue factors had nonsignificant, yet negative correlations - well being (r = -.154), emotionality (r = -.114), sociability (r = -.069), suggesting that persons scoring lower in emotional intelligence scored higher in the number of lies told in a week. When controlled for gender, well being was significant

at .05 level (r = -.169) Gender and age differences between TEIQue factors remained nonsignificant.

The difference between the number of lies told in a week and 5 age groups of respondents was also nonsignificant, F(4,1) = .829, p > .05, however Figure 1 shows that the younger respondents admitted to say more lies than the older respondents.

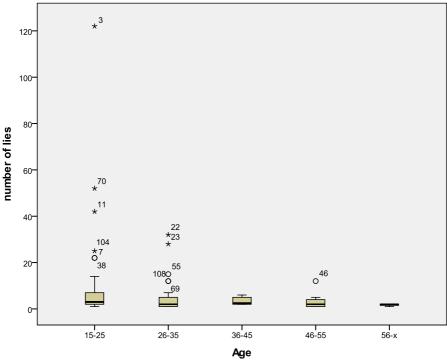


Figure 1: Estimated number of lies for 5 age groups of respondents

The most often mentioned reasons to lie (Table 2) correspond with the reasons found in foreign literature. Women would lie significantly more often in 2 cases - to 'Acquire a reward', F(110,1) = 9.243, p < .01, and to 'Gain admiration from others', F(110,1) = 4.293, p < .05. This order is independent on the relationship between the respondent and a potentionally deceived person and also on the global trait of emotional intelligence. Among other factors, there is a significant relationship between the well-being score and the reason: 'Keeping a secret', F(99, 12) = 2.058, p < .05, and between the estimated number of lies told in a week and the reason 'Gaining admiration from others', F(93,18) = 2.193, p < .01.

Table 2. Most often mentioned reasons to lie

Reason	Ν	Percents
Keeping a secret	101	90%

Protecting someone against punishment	96	86%
Getting out of an unpleasant situation		66%
Avoiding embarrassment		49%
Avoiding punishment		39%
Gaining admiration from others	12	11%
Acquiring a reward	9	8%
Winning control over others	5	4%

A significant relationship between gender and evaluation of lies as prosocial or serious also was not confirmed. Evaluating lies as prosocial or serious was independent on the global trait of emotional intelligence with one exception – respondents with higher global trait score would more often lie to their friends about their looks, F(9,1) = 4.760, p < .05.

Deceiving strangers is easier than deceiving friends or intimate partners for 94.6 % of respondents. The effect of acquaintance is evident here and is independent on gender, F(110,1) = 1.278, p > .05. Yet respondents admitted to lie to strangers and friends or partners with the same frequency, independent on gender or age of respondents. For 75 % of respondents, lying to children is easier than lying to adults. This result is independent on gender, F(110,1) = .127, p > .05.

Correlations between TEIQue scores, relationship between liar and deceived person and the age of deceived person were nonsignificant except for lying to children/adults. Respondents, who found it easier to lie to children, scored higher in global trait (r = .227, p < .01), well-being (r = .197, p < .05), emotionality (r = .156, p < .05) and sociability (r = .187, p < .05).

A stepwise regression analysis showed that among the main variables followed in the research, gender, age and global trait EI together could explain 11.8 % of variance of the frequency of lying, which is significant, F(108,3) = 4.830, p < .01. Other variables were excluded from the analysis.

Among other factors followed in the research, the sense of humour and similar hobbies were significant for both global trait of emotional intelligence and the estimated number of lies at the .01 level, suggesting, that respondents with higher emotional intelligence, who admit to say more lies in a week, look for the sense of humour and similar hobbies when making new relationships.

Analysis of variance revealed that the relationship between liars and most often deceived persons had no significant effect on the number of lies told in one week and was also in nonsignificant relationship with the global trait of emotional intelligence.

5 Conclusions

Current results suggest there are culturally determined differences in certain aspects of lying behavior. This conclusion is supported by foreign studies, although they usually focus on moral development and moral evaluation of lying rather than on lying itself and they mostly compare respondents from China and Canada or the USA. It would be interesting to prove on a larger population sample whether the differences in behavior of liars are typical for certain population groups or whether lying behavior of respondents from middle European countries has similar characteristics. As we have seen, various nations give dissimilar importance to lying and their concepts of lying may differ. It might be beneficial to study the general concepts of lying in the Czech Republic.

Unlike the respondents in foreign researches Czech men admit to say significantly more lies than Czech women, however the frequency of lying does not depend on age, which corresponds with the previous studies. Similar to the results in some of the foreign researches Czech participants also consider lying to strangers easier than lying to their intimate partners or friends. However, there are studies stating the opposite conclusions. Unlike in the foreign studies, it was not confirmed that women said more others-oriented lies than men, moreover women were more willing to lie to gain benefits as admiration or reward. It also seems that gaining admiration is important factor in the frequency of lying. The estimated number of lies negatively correlated with the emotional intelligence scores, suggesting that respondents with lower emotional intelligence admit to say less lies and also tend to lie to adults more often than to children. Contrary to foreign researches the Czech respondents consider lying to children easier than lying to adults.

It also seems that some lies are told to help us maintain our emotional states, as suggested by the relationship between keeping a secret and maintaining higher wellbeing.

The number of participants and the distribution of their characteristics make general conclusions problematic. Most of the participants belong to the group of 15 - 25 years of age. In the future research it would be important to balance the representation of different age groups. The men to women ratio is different to the ratio in the general population and collecting data from more male respondents could bring new answers related to the gender aspects of lying behavior. The total number of participants leads to a non-representative population sample; the current conclusions should be used as a guideline for choosing new research directions or areas and for eliminating methodological problems.

Although the inclusion of participants of various age, professions and from various areas of the Czech Republic secured certain diversity, it would be necessary to focus on the composition of the examined group and consider other possibilities of spreading the questionnaire in the following phases of the research.

The participants' responses can also be influenced by their subjective estimations, given by the character of particular questions. More objective measuring methods as short-term diary recordings could lead to more precise results enabling a comparison of various types of lies and their motivation. This method would be time-consuming and it would be harder to obtain a comprehensive number of respondents. Completing the questionnaire takes a little time and respondents can fulfill it anytime or

anywhere. Advantages of the survey should be viewed together with the disadvantages as absence of a feed-back or the possibility of giving false information, over- or underestimating certain responses.

Understanding the problem can be beneficial in pedagogical psychology for choosing an intervention for excessive lying, in the family therapy for analyzing motivational variables affecting the lying behavior, or in the forensic psychology for analyzing the statements of crime suspects, witnesses or victims. The connection between deception and emotional intelligence might be beneficial in criminal profiling. The comparison of the Czech Republic with other countries could make an interesting contribution of the intercultural psychology to understanding the interpersonal differences induced by the socialization of an individual in a certain cultural-historical society and also understanding the differences on intrapersonal level.

References

Apicella, C.L., Dreber, A., Campbell, B., Gray, P.B., Hoffman, M., & Little, A.C. (2008). Testosterone and financial risk preferences. *Evolution and human behavior*, 29, 6, 384-390.

Aune, R.K., & Waters, L.L. (1994). Cultural differences in deception: Motivations to deceive in Samoans and North Americans. International Journal of. *Intercultural Relations*, 19, 159-172.

Bok, S. (1982). Secrets. Pantheon, New York.

Boon, S.D., & McLeod, B.A. (2001). Deception in romantic relationships: Subjective estimates of success at deceiving and attitudes toward deception. *Journal of Social and Personal Relationships*, 18, 463-476.

Bussey, K. (1999). Children's categorization and evaluation of different types of lies and truths. *Child Development*, 70, 1338–1347.

Caspi, A., & Gorsky, P. (2006). Online deception: Prevalence, motivation and emotion. *Cyberpsychology & Behavior*, 9, 1, 54-59.

Chen, Y., Katuscak, P., & Ozdenoren, E. (2005). *Why can't a woman bid more like a man* Prague: CERGE-EI.

Cole, T., Leets, L., & Bradac, J.J (2002). Deceptive message processing: The role of attachment style and verbal intimacy markers in deceptive message judgments. *Communication Studies*, 53, 74-89.

Croson, R., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, 47, 2, 1-27.

DePaulo, B.M., & Bell, K.L. (1996). Truth and investment: Lies are told to those who care. *Journal of Personality and Social Psychology*, 71, 703–716.

DePaulo, B.M., & Kashy, D.A. (1998). Everyday lies in close and casual relationships. *International Relations and Group Processes*, 74, 1, 63-79.

DePaulo, B.M., Kashy, D.A., Kirkendol, S.E., Wyer, M.M., & Epstein, J.A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, 70, 979-995. Ekman, P. (1992). Telling lies: Clues to deceit in the marketplace, politics, and marriage. WW Norton and Company, New York.

Ennis, E., Vrij, A., & Chance, C. (2008). Individual differences and lying in everyday life. *Journal of Social and Personal Relationships*, 25, 1, 105-118.

Erez, A, Johnson, D.E., & Judge, T.A. (1995). Self-deception as a mediator of the relationship between dispositions and subjective well-being. *Personality and individual deifferences*, 19, 5, 597-612.

Feldman, R.S., Tomasian, J.C., & Coats, E.J. (1999). Nonverbal deception abilities and adolescents' social competence: Adolescents with higher social skills are better liars. *Journal of Nonverbal Behavior*, 23, 3, 237-249.

Ford, Ch.V. (1999). Lies! Lies!! Lies!!!: The psychology of deceit. American Psychiatric Publishing, New York.

Fu, G., Xu, F., Cameron, C.A., Heyman, G., & Lee, K. (2007). Cross-cultural differences in children's choices, categorizations, and evaluations of truths and lies. *Developmental Psychology*, 43, 278–293.

Fu, G., Lee, K., Cameron, C. A., & Xu, F. (2001). Chinese and Canadian Adults' Categorization and Evaluation of Lie- and Truth-Telling about Pro- and Anti-Social Behaviors. *Journal of Cross Cultural Psychology*, 32, 740-747.

Granhag, P.A., & Vrij, A. (2005). Deception detection. In: M. Brewer, & D.W. Kipling (Eds.), *Psychology and law: an empirical perspective* (pp. 43-92), Guilford Press, New York.

Gudjonsson, G.H. (1994). Psychological evidence in court. In: S.J.E. Lindsay, & E.P. Graham (Eds.), *The handbook of clinical adult psychology* (pp. 705-721), Routledge, New York.

Hancock, J. (2007). Digital Deception: When, where, and how people lie online. In: K. McKenna, T. Postmes, U. Reips, & A. Joinson (Eds.), *Oxford handbook of internet psychology* (pp. 287-301), Oxford University Press, Oxford.

Hancock, J.T., Thom-Santelli, J., & Ritchie, T. (2004). Deception and design: The impact of communication technologies on lying behavior. *Proceedings, Conference on Computer Human Interaction*, 6, 130-136.

Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, 435, 673–676.

Lee, K., Xu, F., Fu, G., Cameron, C.A., & Chen, S. (2001). Taiwan and Mainland Chinese and Canadian children's categorization and evaluation of lie- and truthtelling: a modesty effect. *British Journal of Developmental Psychology*, 19, 525–542. Lewis, M., & Saarni, C. (1993). Lying and deception in everyday life. The Guilford Press, New York.

Mancas, M.S. (2007). The New French Autobiography: Considerations on an Aestethics of Lying. In: J. Mecke (Ed.), *Cultures of lying: theories and practice of lying in society, literature, and film* (pp. 317–324), New York: Galda & Wilch. Mannion, G. (2009) *The effect of self-deception on the relationship between trait emotional intelligence and wellbeing*. Darwin: 44th Australian Psychological Society Annual Conference.

Mealy, M. Stephan, W., & Urrutia, I.C. (2007). The acceptability of lies in the United States and Ecuador. *International Journal of Intercultural Relations*, 31, 6, 689-702. Metts, S. (1989). An exploratory investigation of deception in close relationships. *Journal of Social and Personal Relationships*, 6, 159-179.

Mynaříková, L. (2012). *Deception as a Form of Communication in the Czech Republic*. Madrid: Identity, Culture and Communication.

Mynaříková, L., & Boukalová, H. (2012). *Non-verbal lie-detection and characteristics of successful human lie-detectors in the Czech Republic*. Cyprus: European Association of Psychology and Law Conference.

Newton, P., Reddy, V., & Bull, R. (2000). Children's everyday deception and performance on false-belief tasks. *British Journal of Developmental Psychology*, 18, 2, 297-317.

Patterson, J., & Kim, P. (1995). The day America told the truth: What people really believe about everything that matters. Prentice Hall, Englewood Cliffs.

Peterson, C. (1996). Deception in intimate relationships. *International Journal of Psychology*, 31, 279-288.

Petrides, K. V. (2009). *Technical manual for the Trait Emotional Intelligence Questionnaires (TEIQue)*. London: London Psychometric Laboratory.

Porter, S., ten Brinke, L., Baker, A., & Wallace, B. (2011). Would I lie to you? "Leakage" in deceptive facial expressions relates to psychopathy and emotional intelligence. *Personality and Individual Differences*. 51, 133-137.

Reddy, V. (2007). Getting back to the rough ground: deception and 'social living'. *Biological Sciences*, 362, 621-637.

Sweet, M.A., Heyman, G.D., Fu, G., & Lee, K. (2010). Are There Limits to Collectivism? Culture and Children's Reasoning About Lying to Conceal a Group Transgression. *Infant Child Development*, 19, 4, 422–442.

Talwar, V., & Lee, K. (2002). Emergence of white-lie telling in children between 3 and 7 years of age. *Merrill-Palmer Quarterly*, 48, 160-181.

Talwar, W., Gordon, H.M., & Lee, K. (2007). Lying in the Elementary School Years: Verbal Deception and Its Relation to Second-Order Belief Understanding. *Developmental Psychology*, 43, 3, 804–810.

Talwar, V., Murphy, S.M., & Lee, K. (2007). White lie-telling in children for politeness purposes. *International Journal of Behavioral Development*, 31, 1, 1-11. Tyler, J.M., Feldman, R.S., & Reichert, A. (2006). The price of deceptive behavior: Disliking and lying to people who lie to us. *Journal of Experimental Social Psychology*, 42, 1, 69-77.

Van den Bergh, B., & Dewitte, S. (2006). Digit ratio (2D:4D) moderates the impact of sexual cues on men's decisions in ultimatum games. *Proceedings of the Royal Society B*, 273, 2091–2095.

Vrij, A. (2007). Credibility assessments in a legal context. In: D. Carson, B. Milne, F. Pakes, K. Shalev, & A. Shawyer (Eds.), *Applying psychology to criminal justice* (pp. 81-96), John Wiley and Sons Ltd., London.

Vrij, A. (2008). Detecting lies and deceit. England: Wiley, Chichester.

Westcott, H.L, Davies, G.M., & Clifford, B.R. (1991). Adult's perceptions of childern's videotaped truthful and deceptive statements. *Children & Society*, 5, 2, 123-135.

Whitty, M.T. (2002). Liar, liar! An examination of how open supportive and honest people are in chat rooms. *Computers in Human Behavior*, 18, 4, 343-352.

Whitty, M.T., & Gavin, J. (2001). Age/sex/location: uncovering the social cues in the development of online relationships. *CyberPsychology & Behavior*, 4, 5, 623-630. Wilson, A.E., Smith, M.D., & Ross, H.S. (2003). The nature and effects of young children's lies. *Social Development*, 12, 21-45.

The Construction and Evolution of the Romantic Relationships in Cyberspace

Floare Chipea, Raluca Sînziana Buhaş

Doctoral Student at the Faculty of Social Sciences, University of Oradea, Romania, raluca_buhas@yahoo.com PhD. Professor at the Faculty of Social Sciences, University of Oradea, Romania, fchipea@yahoo.com

Abstract. The contemporary society is characterized by a profound implication of the Internet in daily life. The study aims to investigate the influence of the Internet on romantic relationships' dynamics. Modern ways of selecting a romantic partner are framed due to cyberspace. Which are the construction patterns of an online romantic? Are the same as those characteristic for the offline context? How does an online relationship develop beyond cyberspace? To obtain a complex view of the subject, the proposed sociological methodology will follow a mixed process. The quantitative method will use the investigation based on an online questionnaire applied to the selected respondents. A semi-structured interview guide will be applied to romantic couples, many of them married. This qualitative technique will outline interesting elements for the couple's story. The fact that the scientific literature focusing on virtual sociology – especially on online relationships, is not very developed in the Eastern Europe, makes a research on this topic very useful and interesting.

Keywords: cyberspace, relationships, partner selection process, computer-mediated-communication

1 Introduction

The 20th and 21st centuries are considered as the beginning of the postmodern era and we cannot relate to them without taking into account an important feature – the influence of technology in everyday life. The society is subject to continuous dynamics, closely linked with the development of the technology, creating new frames of social manifestation. The modern technology of the Internet came to mark and transform almost every facet of our life. The human relationships' aspect is not excluded from this rule. Cyberspace allows millions of people to establish and maintain relations, even if they are at large geographical distances. Due to virtual space's features, human connections that exist within it have specific evolution patterns. Thus, sociologists have recently begun to study virtual reality and its implications on society. The specialized literature has been enriched with new subjects from cyber-sociology or sociology of the Internet.

Traditionally, individuals choose their romantic partner through networks of friends, relatives, at work or in public socializing places. Due to cyberspace, new ways of choosing the marital partner arose. Online dating sites and online social networks are rapidly imposing on the "marital market" (Becker, 1994, p.11). This method of socializing becomes more and more common among modern society's members. It is considered to be very convenient. Some experts argue that an effect of this new type of approach can be seen in the decreasing intensity of the spatial proximity phenomenon (Ilut, 2005).

Which are online relationships' formation patterns? What is the profile of the social actors involved? The proposed research aims to follow the exploratory sociological analysis' line, as well as the descriptive analysis.

2 Literature review

The contemporary romantic couple is very different to the traditional one. The modern human mate selection process meets two directions of transformation: is performed at individual level - everyone chooses their romantic partner according to personal values; is mainly based on individual's emotional-affective needs

The scientific literature frames two main theoretical directions regarding human mate selection: *the evolutionary perspective* and *the psycho-social perspective* (Kapadia, 2001). The first one emphasizes that the romantic partner is selected based on an unconscious mechanism that values the genetic dowry. The second perspective focuses on more pragmatic elements: everybody seeks a partner with an equivalent value to its own.

B. Murstein developed *the marital choices model as a multifactorial process* (Murstein cited in Ilut, 2005). It implies three phases before the final establishment of the couple: the focus on mutual stimulus (e.g. impression about physical appearance); the value compatibility level; the role's needs compatibility.

Beside socio-demographic determinants, the psychosocial mechanisms influence human mate selection process. Some of the psychological factors involved are (Iluţ, 2005): the physical attractiveness, the excitability transfer, the "Romeo and Juliet" effect, the "hard to get" effect, the similarity and complementarity process.

In scientific literature, two theoretical directions regarding computermediated communication are framed: *positive theories and negative theories*.

Negative theories (Whitty, 2008) sustain that online relationships can be a negative and impersonal experience that limits the time for "healthy" offline relationships.

1. <u>The social presence and gesture's theory</u> (Whitty, 2008) - argues that online relationships are more impersonal than the offline ones due to the limitations of the

non-verbal communication in cyberspace. A decline in social presence is generated (Hiltz, Johnson & Turoff, 1986; Sproull & Kiesler, 1986).

2. <u>The 'media abundance' theory</u> (Whitty, 2008) - compares online with offline communication, highlighting the limits of the first one. Direct communication is considered to be the most offering channel in transmitting the information, opposed to online transmitted information.

Positive theories are supported among others by Suler (2004) and Joinson (2001) and state that individuals can be more honest and open with each other when they develop a conversation in cyberspace.

1. <u>"The disinhibition effect"</u> (Whitty, 2008). Using the mask of anonymity offered by cyberspace, the individual has the feeling of a greater freedom in expressing emotions. Suler (2004) distinguishes *"the benign disinhibition"* when people reveal their fears, personal secrets and feelings and *"the toxic disinhibition"* when, under anonymity people can choose to express their negative feelings and emotions.

2. <u>The hyper-personality theory</u>. Developed by Whalter and his collaborators (2001), it sustains that computer-mediated romantic relationships can develop faster and have a higher sense of intimacy than the one showed by offline relationships. Online partners can give more social and emotional support.

3. <u>The presentation of `the true self</u>. McKenna and his colleagues (2002) suggest that Internet is a suitable context in shaping human relationships, especially for those who declared themselves to be lonely and had high social anxiety. The Internet is considered a safer place than the actual reality when presenting `the true self'.

4. <u>'The play in cyberspace'</u>- developed by Whitty (Whitty & Carr, 2006) argues that Internet can be considered a playground for individuals, because they can experience the presentation of their own self. An online relationship can be more easily to manage, because the Internet is a `not so real` world.

3 Proposed Methodology

The interest for the present study developed some major research *objectives*:

1. Outlining the socio-demographic profile of individuals who develop online romantic relationships

2. Identifying the developmental stages of online romances` formation

3. Analysing online romance's dynamics in offline reality (aspects regarding the period of time passed until the online relationship develops offline, the offline lifetime of this type of romance etc.)

According to the objectives mentioned above the following *research hypothesis* were framed:

- Cyberspace, as a modern method to select the marital partner is accepted by young adults and rejected by the middle-aged and elderly people
- The developmental stages of romantic relationships initiated on cyberspace differ from those that characterise offline relationships the communicational aspect is more emphasized
- The online initiated romantic relationships are more intimate and develop faster than offline ones.

• The social institutions (family, community) have no influence in the online process of spouse selection.

To analyse more closely the proposed research topic, a mixed methodological approach is necessary. *The quantitative technique* will follow the sociological inquiry method. An online questionnaire will be applied to samples consisting of Romanian users over 18 years who are active on dating sites and different social networks. Due to theme's novelty, the sample will be random. Its uncertain representativeness represents one limit of the study. The results will be statistically analysed in order to test the hypothesis.

Interesting data might be offered also by an international comparison on samples of subjects who used cyberspace in order to meet their life partner and who developed their relationship online. This comparison will be done by realizing a *secondary statistical analysis* on data gathered from European surveys and not only.

Qualitatively speaking semi-structured interviews will be used. They will be applied directly to couples which have begun their relationships in cyberspace and materialized them in offline reality, even by marriage. 10 couples that were identified up to now will constitute the *case studies* for the research. *The focus-group method* will be used to frame Romanian respondents' perception on the acceptance of cyberspace as a modern way to socialize and to choose a romantic partner. A semi-structured guideline interview will be applied to 3 groups of respondents: young adults, middle-aged people and elderly.

4 Conclusions

The research aims to contribute to the scientific study of cyberspace and its influence on human relationships, especially on romantic ones. Even if the study cannot be considered exhaustive, it may provide an interesting and real opening in this domain, taking into consideration that few researches have been conducted in this sense. Peculiarities of cyberspace frame new sociability patterns and provide an overview of society's future.

The study presents some methodological limits that could be surmounted by future analyses: new sociological methods adapted to cyber-reality can be used; the results of the research cannot be extended to other national and cultural contexts.

The online relationships involve various forms of manifestation "enriching the human sociability spectrum" (Nadolu, 2004, p 242). The control that users have on the online presentation of their image, the possibility of direct communication without "the interference" of non-verbal cues, are specific features of cyber-reality. It is what makes the online communication distinct, enriching the human relationships' spectrum.

References

Becker, G. (1994) Human Behaviour – An Economic Approach, All Publishing House, Bucharest

Hiltz, S., Johnson, M. & Turoff, M. (1986) Experiments in Group Decision Making: Communication Process and Outcome in Face-To-Face Versus Computerized Conferences. *Human Communication Research*, Vol. 13, 225-252

Iluţ, P. (2005) Socio-psychology and anthropology of the family, Polirom Publishing House, Iaşi

Joinson, A. N. (2004) Self-esteem, interpersonal risk, and preference for e-mail to face-to-face communication. *CyberPsychology and Behavior*, Vol. 7, no. 4, 479-485

Kapadia, S. (2001) Psychology of Partner Selection across Culture, in Smelser, N. J. & Baltes, P. B. (eds), *International Encyclopedia of the Social & Behavioral Sciences*, Elsevier, Amsterdam-New York, 11090-11092

McKenna, K. Y. A., Green, A. S., & Gleason, M. E. J. (2002) Relationship formation on the Internet: What's the big attraction?. *Journal of Social Issues*, Vol. 58, 9-31

Nadolu, B. (2004) The sociology of virtual spaces, Eurostampa Publishing House, Timişoara

Sproull, L. & Kiesler, S. (1986) Reducing social context cues: Electronic mail in organizational communication. *Management Science*, Vol. 32, 1492-1512

Suler, J. (2004) The online disinhibition effect. *CyberPsychology and Behavior*, Vol.7, 321-326

Whitty, M. T. (2004) Cyber-flirting: An examination of men's and women's flirting behaviour both offline and on the internet. *Behaviour Change*, Vol. 21, no. 2, 115-126

Whitty, M. T. (2008) Liberating or debilitating? An examination of romantic relationships, sexual relationships and friendships on the Net. *Computers in Human Behavior*, Vol. 24, 1837-1850

Whitty, M. T. & Carr, A. N. (2006) *Cyberspace romance: The psychology of online relationships,* Palgrave Macmillan, Basingstoke

Acknowledgement¹

¹ This work was partially supported by the strategic grant POSDRU/88/1.5/S/53501, Project ID53501 (2009), co-financed by the European Social Fund-Investing in People, within the Sectorial Operational Programme Human Resources Development 2007-2013.

Heart Rate Variability during Decision-Making in Children Considering Piaget's Theory of Cognitive Development

Mária Túnyiová¹, Ivan Sarmány-Schuller¹

¹ Institute of Experimental Psychology, Slovak Academy of Sciences, Dúbravská cesta 9, 813 64 Bratislava, Slovak Republic maria.tunyiova@savba.sk

Abstract. The aim of this study is to investigate the somatic markers in the process of decision-making within the Iowa Gambling Task in 6 - 13 year old children. In accordance with Piaget's theory of cognitive development, children will be divided into three groups – preoperational stage, concrete operational stage and formal operational stage – based on their success in Piaget's experiments - conservation tasks and pendulum task. Modified Iowa Gambling Task (IGT) will be played and, simultaneously, heart rate variability will be measured. We hypothesize that children in the preoperational stage and concrete operational stage will prefer decks with higher frequency of rewards (B, D) with small regard to gained or lost money and, additionally, we suppose bigger anticipatory decrease of heart rate before selecting a card from decks with high frequency of punishments (A, C). We hypothesize that children in formal operational stage will prefer advantageous decks (C, D) and, additionally, we suppose bigger anticipatory decrease of heart rate before selecting cards from disadvantageous decks (A, B).

Keywords: *decision making, Iowa Gambling Task, somatic markers, heart rate variability, Piaget*

1 Somatic marker hypothesis and cognitive development

Somatic markers are somatic signals based on past experiences, when the body learned that a similar incident led to pleasant or unpleasant feelings. The ambition of empirical research on somatic markers led to the development of the Iowa Gambling Task (IGT) (Bechara, Damasio, Damasio & Anderson, 1994), which included

rewards, losses and uncertainty, and also simulated decision-making in real life. Test materials consist of four decks of cards, A, B, C, D, with fifty cards in each deck. Monetary rewards and losses are labeled on the down side of each card. The gains and losses vary as displayed in Table 1. Decks A and B are disadvantageous, decks C and D are advantageous in terms of expected value of net total score. Participants are asked to select cards in a way to maximize their profit. Anticipatory hearth rate slows down before disadvantageous relative to advantageous option (Crone, Somsen, Van Beek & Van der Molen, 2004). The theory of somatic markers implies importance of feelings or intuition in decision-making. Sarmány-Schuller (2011) found out that people, who prefer analytical strategies.

	Bad Decks		Good	Good Decks	
	A	В	С	D	
gain/deck	\$100	\$100	\$50	\$50	
loss/10 cards	\$1250	\$1250	\$250	\$250	
net/10 cards	-\$250	-\$250	\$250	\$250	
losing card/10 cards	5	1	5	1	

 Table 1: Decks of the Iowa Gambling Task

Crone, Bungem, Latenstein & Van der Molen (2005) say that children up to approximately 12 years of age do not have somatic markers and they concluded that this results in failure of children in IGT. The research of Crone, Vendel & Van der Molen (2003) showed that children older than 12 years were able to be successful in IGT and the performance positively correlated with age. Hooper, Luciana, Conklin & Yarger (2004) studied adolescents' performance in IGT. The result of their study confirms that children are able to understand IGT and their success in game is positively correlated with age. The aim of our study is to provide possible explanation of the success of children older than 12 years in IGT by using the theoretical framework of Piaget's theory of cognitive development.

1.1 Piaget's theory of cognitive development and IGT

Preoperational stage (2-7 years) of cognitive development is characterized by centration, the capacity to focus on the only one aspect of a problem at a time (Piaget & Inhelder, 1958). We suppose that children in preoperational stage will focus (or will be centrated) on frequency of rewards with little regard to amount of gained or lost money in IGT. In this case we expect higher preference for decks B and D with high frequency of rewards and, simultaneously, we expect bigger slowing of heart rate before selecting cards from decks A and C in comparison to selecting cards from deck B and D.

732

The thinking of children in concrete operational stage (7-11 years) is decentrated and they are able of conservation, the ability to understand that physical attributes of an object remain unchanged even though their appearance has changed (Piaget & Inhelder, 1958). In concrete operational stage children are able to perform operations such as calculating and classification; they understand logical relations between concrete objects and events. Because the thinking is still based on ideas about concrete objects, we suppose that children in concrete operational stage will not be able to recognize all aspects of each deck and relations between decks. We expect that children in concrete operational stage will still focus mainly on frequency of rewards with little regard to the amount of gained or lost money. They will prefer decks B and D and we expect bigger slowing of heart rate before selecting cards from decks A and C in comparison to selecting cards from decks B and D.

Children in formal operational stage (11 years and more) are able to think logically and process highly abstract ideas (Piaget & Inhelder, 1958). We suppose that children in formal operational stage will be able to recognize all aspects of cards in IGT and relations between them and in this case they will prefer selecting cards from decks C and D and at the same time, we expect bigger slowing of heart rate before selecting cards from decks A and B in comparison to selecting cards from decks C and D.

2 Method

2.1 Participants

Approximately 20 participants in each stage (preoperational, concrete operational and formal operational stage), together 60 participants, consisting of 50% boys and 50% girls, age 6-13 years, will participate in our research. We decided to study children in the second half of the school year, because children in first grade already know numbers and have basic mathematic skills (they understand plus and minus). The bottom limit, 6 years of age, is given because most children should be in preoperational stage. The upper limit, 13 years of age, is given because most children in this age are already in formal operational stage. Children participating in the research will be from public school, without any known psychological or somatic disease. All parents will be asked for permission to allow research with their child.

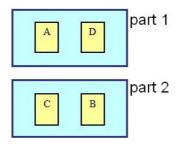
2.2 Materials and procedure

Participants will be divided into three stages of cognitive development based on their performance in Piaget's experiments. At first, children will solve the conservation tasks – the water level experiment and conservation of number task. If a child succeeds, he/she will be asked to solve the pendulum task. In case the child is not able to solve any of the tasks, he/she will be considered to be in preoperational stage. If a child is be able to solve conservation tasks, but not the pendulum task, he/she will be considered to be in concrete operational stage. If a child is able to solve all three tasks, he/she will be considered to be in formal operational stage. We decided to modify the IGT. At first, we will change currency from dollars to euros, because euro is used in our country and children will understand the value better in euros than in dollars, and eventually use them in everyday life too. We decided to use one tenth of the original amount of money, because we suppose it will be easier to understand the value of money in the game.

We will divide IGT into two parts, as displayed in Figure 1. In the first part of the game, participants will select a card 100 times, either from disadvantageous deck A (on the left side) or advantageous deck D (on the right side). After playing the first part, participants will play the second part, where they will select a card 100 times, either from advantageous deck C (on the left side) or disadvantageous deck B (on the right side). Participants will not get information about which deck is advantageous. Each studied trial will consist of 20 selections (5 trials in the first part and 5 trials in the second part of the game).

We will measure heart rate variability with photopletysmograph and we will focus on the magnitude of heart rate one second before selecting a card. From the measured data we will identify the main anticipatory heart rate of each deck. We will also study the main anticipatory heart rate of each trial of each deck. Score will be calculated for each trial and net score will be stated as D minus A, respectively C minus B.

Figure 1: Arrangement of decks in the first and second part of the game



3 Expected contribution

However, Crone, Bungem, Latenstein & Van der Molen (2005) say that because children younger than 12 years are not able to recognize advantageous decks, they do not have somatic markers. We expect to confirm that children do have somatic markers, but with respect to cognitive development, they have different preferences and make different decisions in comparison to healthy adults.

References

- Bechara, A., Damasio, A. R., Damasio, H., Anderson, S. (1994) Insensivity to future consequences following damage to human prefrontal cortex. *Cognition*. Vol, 50, pages 7-15.
- Bechara, A., Tranel, D., Damasio, H., Damasio, A. R. (1996) Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral Cortex*. Vol, 6, pages 215-225.
- Crone, E: A., Bunge, S. A., Latenstein, H., Van der Molen, M. W. (2005) Characterization of children's decision making: sensitivity to punishment frequency, not task complexity. *Child Neuropsychology*. Vol, 11, pages 245-263.
- Crone, E. A., Somsen, R. J. M., Van Beek, B., Van der Molen, M. W. (2004) Heart rate and skin conductance analysis of antecedents and consequences of decision making. *Psychophysiology*. Vol, 41, pages 531-540.
- Crone, E. A., Vendel, I., Van der Molen, M. W. (2003) Decision-making in disinhibited adolescents and adults: insensivity to future consequences or driven by immediate reward? *Personality and Individual Differences*. Vol, 34, pages 1-17.
- Hooper, C. J., Luciana, M., Conklin, H. M., Yarger, R. S. (2004) Adolescents' performance on the Iowa Gambling Task: Implications for the development of decision making and ventromedial prefrontal cortex. *Developmental Psychology*. Vol, 40, pages 1148-1158.
- Piaget, J., Inhelder, B. (1958) *The Growth of Logical Thinking, from Childhood to Adolescence*, Basic Books, inc., New York.
- Sarmány-Schuller, I. (2011) Vplyv kognitívneho štýlu a procesov rozhodovania na múdrosť (Impact of cognitive style and decision-making processes on wisdom). in: Ruisel, I., Ruiselová, Z (eds), Múdrosť v kontexte osobnosti: problémy, fakty, otázky, Ústav experimentálnej psychológie SAV, Bratislava, pages 165-189.

Substance misuse treatment process evaluation: A dimensional trait-based conceptualization

Fivos E. Papamalis*, Efrosini Kalyva*, and Petra S. Meier^

*South East European Research Centre, Greece ^School of Health and Related Research University of Sheffield, UK Email: fpapamalis@seerc.org

Abstract. Despite evidence of the association of certain traits with drug use initiation, misuse and relapse, to date limited efforts have been made to identify their role in treatment process. In this context, the complex relationship among clients' enduring personality traits and their phenotypic expression in treatment is of particular concern. In line with the current developments of the dimensional based conceptualizations and progressive methods of discriminating adaptive and maladaptive personality traits, this study examines whether and to what extent, different personality dimensions contribute to or hinder individuals' treatment engagement. Potential influence of organizational factors on clients' engagement is also explored. The process of therapy will be examined in a naturalistic setting in the main substance misuse treatment organizations in Greece. Longitudinal multi-site design is utilized to explore the relationship between study variables in different treatment settings and examine whether there are baseline differences by using pre- and during process measures. The sample involves service users and the clinical staff. The assessment battery involves CEST Intake/CEST (Simpson, 2001,2005); SIPP-118 (Verheul et al., 2008); TPQue (Tsaousis, 2002) and ORC (Lehman et al., 2002). The importance of disentangling the major overlap among personality traits and disorders, the discrimination of traits from characteristic adaptations and their reciprocal relationship with contextual environment is highlighted. Associations among characteristic adaptations, personality traits and engagement would imply a broader conceptual framework in which engagement modifications are viewed in the context of this interaction. Delineating the role of personality dimensions within treatment process may contribute

to the identification of individual attrition vulnerabilities and individualized interventions.

Keywords: treatment engagement, process evaluation, personality traits, characteristic adaptations, co morbidity

I Introduction

One of the core questions that have been put forward in the scientific community refers to the effectiveness of treatment services and service users' social reintegration. Empirical studies have consistently demonstrated that drug treatment is effective and produces significant measurable changes in drug use and other behaviours, compared with no treatment or a pre-treatment status (Simpson et., 1997; Simpson, Joe, & Broome, 2002). However, early drop-out represents a wide spread problem in substance misuse treatment (Bargagli et al., 2006; Darke et al., 2011; Mulder et al., 2009) and is associated with a number of negative implications for individuals, treatment providers and broader community. Data worldwide indicate increased rates of drop out with over 60% on individuals leaving treatment early on (Cournoyer et al., 2007; Darke et al., 2011; Jones et al., 2009; Mulder et al., 2009).

Although there is an international trend of science-based interventions in order to increase treatment effectiveness, there is a necessity to achieve the appropriate balance between the ideals of "best practice" and the particular needs of each individual. In this line, modern initiatives support that evaluation represents a step forward to assure quality and raise standards of care (EMCDDA, 2008). Despite perceived homogeneity related to epiphenomenal consequences of drug use, individuals who misuse substances present quite different problems that vary with development status, predisposed tendencies and socio-cultural adjustment. Current shift in research evolved towards the recognition that a number of personality traits commonly observed in drug users do not necessarily reflect pathological states (Livesley, 2007). The scientific debate on understanding psychopathology and formulating treatment guidelines based on multileveled conceptualization of personality dimension has lately received increased support. Thus, further investigation of the role of personality in treatment is of vital importance to solidify our knowledge in regards to potential clinical improvements that could be achieved if therapeutic interventions were tailored according to individual differences.

II Literature Review

2.1Treatment engagement and co morbidity

Several studies in general psychotherapy and substance misuse indicate that clients' active engagement in therapy has been consistently associated with increased tenure

in treatment, reduction of drug use and improved post treatment outcomes (Crits-Cristoph & Connolly, 2003; Fiorentine & Anglin, 1997; Hser et al., 2001; Simpson, et al., 1997; Simpson, 2004; Simpson & Danserau, 2008). Despite the lack of conceptual differentiation of engagement and related constructs such as motivation, participation and alliance, research provides significant evidence of their role in the therapeutic process. In contrast, failure to engage or premature termination has been associated with a variety of client-related factors, including demographics such as socioeconomic status, employment, education, gender (e.g. Claus & Kindleberger, 2002), criminal history, chronic drug use, previous treatment experiences (e.g. Stark, 1992) and treatment expectations (e.g. Fiorentine, Nakashima, & Anglin, 1999). Moreover, clients' psychiatric severity (Broome, Flynn, & Simpson, 1999; McLellan, Luborsky, & Woody, 1983; Claus& Kindleberger, 2002; Sweet & Noones, 1989; Herbeck et al., 2005) and specifically antisocial and borderline personality disorders or traits associated with these disorders (Clarkin & Levy, 2004; Meier & Barrowclough, 2009) have been related with increased vulnerability to premature drop-out. Similarly, studies demonstrate that individuals with undetected additional disorder or unmet needs during treatment are more likely to drop out (e.g. Schulte et al., 2010).

In this context, high comorbidity rates are of particular concern considering that approximately more 50% to 70 % of individuals in substance misuse treatment have at least one additional psychiatric diagnosis and particular personality disorder (e.g. Brown et al., 2011; Ball et al., 2006). These estimations indicate that dual diagnosis is the norm, rather than the exception among substance abusing individuals, which is quite concerning as there is a lack of empirical support for efficacious interventions. As a result, co-occurring psychiatric disorders and their clinical management represent a key complicating factor in treatment as it is often associated with lack of treatment engagement, non-compliance, poor retention, and greater risk for relapse (Brown et al., 2011; see for reviews Kreyenbuhletal., 2009; O'Brien, Fahmy, & Singh, 2009).

2.2 Dimensional trait-based conceptualization

The wide spread use of the term "co occurring disorders" includes a multiplicity of diagnostic categories that covers all types of substance use and mental health disorders, without distinguishing the levels of severity and intensity of these conditions. Moreover, the extensive overlap among diagnostic categories brings into question its usefulness in treatment planning process (Flynn & Brown, 2008). Current research evolved towards the recognition that personality traits commonly observed in drug users do not necessarily reflect pathological states, separating substance use from personality pathology (e.g. Livesley, 2007). Considering the high co-morbidity of substance use and personality disorders, which significantly impedes treatment progress (van den Bosch & Verheul, 2007) it is necessary to consider that personality disorder is being over-diagnosed due to the overlapping symptomatology. This may be related to the presence of active Axis I substance use disorder, which alters cognitive, emotional and behavioural functioning and may result in exacerbation of

symptoms, mimic personality pathology and complicate differential diagnosis (Ball, 2005; Dawson et al., 2005; Jentsch & Taylor, 1999).

Moreover, diagnostic indicators of ruling out personality disorder are based on the behavioural expression and extremity of certain personality traits. Nevertheless, trait elevation or extremity does not represent determining or sufficient criteria for diagnosing disorder (Livesley & Jang, 2000; Verheul, 2005). This suggests that discrimination of trait ambiguities offers important clinical applications as it may decrease fluctuations of personality pathology with substance misuse. That is, clinical significance involves harmful dysfunction in the adaptive functions of the dynamic personality system, and not just statistical deviance on personality traits (e.g. Verheul, 2005).

In this line, several prominent authors including DSM-V and *ICD*-10 task forces have questioned the fundamental appropriateness of the categorical model and suggest it should be reconceptualised by including progressive methods of dimensional assessment and discriminating adaptive and maladaptive personality traits (Flynn & Brown, 2008; Jackim, 2005; Krueger et al. 2007; Samuel et al., 2010). This approach may provide more reliable scores and reveal valuable information regarding lower-order traits and symptoms (Lowe & Widiger, 2009). Thus, disentangling traits from disorders based on a continuum of intensity and severity indicates the clinical utility of the dimensional approach, as it may improve individualized assessments, enhance treatment specificity and facilitate appropriate personality matching interventions.

Overall, it appears that client characteristics may have different patterns of influence on diverse segments of treatment engagement. Despite substantial advancement offered by the multi-analytic techniques, there is a need for further investigation of factors that contribute to or impede engagement in treatment.

While client pre-treatment characteristics (demographics, mental health problems, drug use) as factors influencing treatment engagement have been extensively researched, knowledge of the role of clients' personality traits and adaptive capacities is lacking. In this context, the complex relationship among clients' personality traits and their phenotypic expression within treatment process is of major importance. This would allow the formulation of invividualized interventions, facilitate individuals' adaptation and allow greater flexibility to respond to the diversified clients' needs. Under this prism, it appears there is a need to explore whether and how personality dimensions are associated with or likely to influence individuals' treatment responses.

III Methodology

3.1 Aims & Conceptualization

The purpose of this study is to examine the extent to which a range of client personality dimensions are related to certain treatment engagement indicators.

Encouraged by previous literature and current trends, the present study will develop an integrative conceptual framework in order to detect and measure possible prognostic indicators that may influence treatment engagement.

Figure1: Visual representation of the hypothesized association of broad personality dimensions, (mal) adaptive personality functioning and treatment engagement indicators

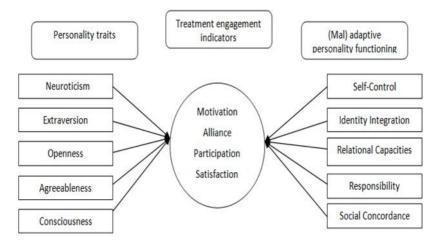
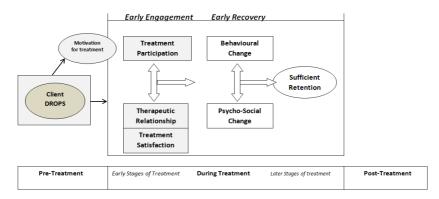


Figure 2: Visual representation of the treatment process model



Note: This figure is modified version of Texas Christian University Treatment Process Model (Simpson, 2001). The coloured areas highlight the scope of the present study, including pre-treatment clients' Dimensional Representation Of Problem Severity (DROPS) that represents clients' personality profile based on broad/lower order traits and characteristic adaptations; and treatment engagement indicators selected for this study. Other stages of treatment (e.g. later stages) along with their elements (e.g. behavioural change) are beyond the current scope.

3.2 Study Design

The process of therapy is examined in a naturalistic setting. The study will test relationships between key variables, examine whether there are baseline differences between clients with different personality characteristics and explore determinants of treatment engagement using pre- and during process measures. A longitudinal multisite design will be utilized to explore the relationship between certain dimensions of clients' personality and engagement in different treatment sites. Quantitative analysis will be used to examine whether personality dimensions predict treatment engagement indicators. The grounds for the utilization of quantitative approach in this process-oriented study are: a) identify patterns and significant associations among major study variables b) examine key predictors among personality dimensions and engagement indicators in the context of substance misuse treatment, b) elucidate and explain the differences and similarities in the levels of engagement according to personality dimensions and c) draw conclusions that may be generalized. Objective measurement based on empirical taxonomy may indicate the influential role of personality in treatment engagement and facilitate the systematic accumulation of findings.

Moreover, in order to overcome measurement-related issues that result from the overlap of traits and adaptations, this study builds upon previous research by integrating personality dimensions including: a) broad descriptive structure of personality and lower facet level traits according to FFM (e.g. Muller, 2008) which are more stable and resistant to change and b) (mal) adaptive personality functioning or characteristic adaptations (Verheul et al., 2008) which may be more malleable to change throughout treatment

3.3 Recruitment of services & participants

The treatment centres involved in the study represent major public treatment facilities that provide comprehensive care for alcohol and substance misuse in Greece. The first type of treatment (Tx1), Therapeutic Community (TC), is a non-profit, non-governmental organisation consisted of nation-wide network of over 70 units. Treatment staff includes counsellors-former service users, social workers and psychologists. TC refers to the traditional therapeutic communities that operate based on highly structured environment. Two outpatient and two inpatient (male/female) treatment community units are participating, with approximately total capacity of 90 beds. The overall number of employed clinical staff is around 16 in both units.

The second type of treatment included in the study (Tx2) refers to psychosocial residential rehabilitation. The main department is based in the Attica Psychiatric Hospital. Although traditionally based on a psychodynamic approach, this centre offers a variety of therapeutic interventions. Treatment staff consists of psychologists, psychotherapists and social workers. Types of therapies provided include long-term inpatient and outpatient treatment, individual and group counselling. In this type of treatment, two male, one mixed inpatient and two outpatient mixed units participate in

the study. The overall capacity is around 50 beds, whereas approximately 12 staff members are employed.

The main phase of therapy is lasting approximately 6-9 months for both treatment types. The treatment entry procedure for both units involves the preparation phase, within which individuals initially receive outpatient support therapy once per week. In both treatment types, after two weeks of individual counselling, clients engage in a more intensive format, including group therapy and individual sessions.

3.4 Recruitment of participants

Participants are being recruited in the outpatient units during the preparation phase, with efforts been made to brief them about the study during the first days since their treatment entry. Recruitment will continue until a sample size of 200 participants who enter in treatment is reached from each treatment type. The recruitment process has been estimated to last approximately six months more than initially planned, that is, the completion is expected in winter 2013. Several reasons have influenced this alteration. Primarily, almost all treatment units had approximately 25% to 40% reduction of the admissions. This may be associated with the new policy of methadone administration in public hospitals and with the general effects of economic downturn related to decreased faith, willingness and motivation to engage in therapeutic process.

3.5 Intake Assessment procedure

Data collection is performed at several time points. The first intake assessment battery is conducted during the initial contact of the individual with the treatment services and includes CEST Intake (CESTI; Simpson, 2001, 2005), SIPP-118 (Verheul et al., 2008), and TPQue (Tsaousis, 2002). As problem severity measured by CEST-I and characteristic adaptations measured by SIPP-118 are malleable to change during treatment, it is considered appropriate to capture these indicators early on. Therefore it has been arranged with treatment providers to administer these questionnaires in the early phase of treatment.

3.6 During treatment assessment procedure

During-process follow-up is performed between during the inpatient treatment, and includes re-administration of the CEST and SIPP-118 to individuals who have participated in the first phase. Finally, the administration process is successfully completed with the staff members from all the units. It is particularly interesting to note that all staff members working in the organizations involved in the study have consent to participate.

Instrument	CEST Intake	CEST	TPQue	SIPP-118	ORC
Author	Simpson, 2001;2005	Simpson, 2001;2005	Tsaousis, 2002	Verheul et al., 2008	Lehman et al., 2002
Constructs measured	Treatment motivation Psychological functioning Social functioning	Treatment Motivation Alliance Participation Satisfaction	Extraversion Neuroticism Openness Consciousness Agreeableness	Self-control Identity integration Relational Capacities Responsibility Social concordance	Staff attributes Organizational Climate
Length in minutes	15	25	30	30	25
Psychometric properties	.86 to .96	.86 to .96	.78 to .89	.69 to .86	.62 to .79
Previously used with SU tx	Yes	Yes	Yes	No	Yes
Greek version available	No	No	Yes	No	No
Hypothesis number	1	2, 3	2, 3	1, 3	3

Table 1: Questionnaires used to assess study variables

IV Conclusion and practical implications

From the clinical research perspective, the study introduces an innovative conceptual framework based on dimensional diagnostic indicators that could contribute to the identification of individual attrition vulnerabilities. The disentanglement of maladaptive traits from diagnostic categories and the clinical significance of characteristic adaptations may assist the identification and modification of potential obstacles of treatment engagement and provide the basis for contemporary individualized personality matched interventions.

Moreover, the downturn of economic recession followed by resource cut-offs in substance misuse treatment services, as well as demands for cost effectiveness, implies the need for efficient and wise spending of public recourses. Treatment process evaluation represents the sole bridge integrating scientific evidence and clinical practice, and may be useful tool or act as compass of reorientation and evidence base application. By integrating two different lines of research on personality and engagement, this study attempts to indentify key ingredients underlying treatment engagement, explore behavioral phenotypes and enhance treatment response specificity according to clients' unique needs.

References

Ball S.A. (2005). Personality traits, problems, and disorders: Clinical applications to substance use disorders. *Journal of Research in Personality*, 39, 84–102.

Ball, S.A, Kathleen, M. Carroll, Ball, M.C., & Rounsaville, B.J. (2006) Reasons for dropout from drug abuse treatment: Symptoms, personality, and motivation. *Addictive Behaviors*, Vol. 31, pp. 320-330.

Bargagli, A. M., Faggiano, F., Amato, L., et al. (2006). VEdeTTE, a longitudinal study on effectiveness of treatments for heroin addiction in Italy: study protocol and characteristics of study population. *Subst Use Misuse*, 41:1861–79.

Brocato, J. & Wagner, E. F. (2008). Predictors of retention in an alternative-to-prison substance abuse treatment program. *Criminal Justice and Behavior*, *35*, 99-119.

Brown, C. H., Bennett, M. E., Li, L., & Bellack, A. S. (2011). Predictors of initiation and engagement in substance abuse treatment among individuals with co-occurring serious mental illness and substance use disorders, *Addictive Behaviors*, *36*, 439–447.

Clarkin, J.F., & Levy, K.N. (2004). The influence of client variables on psychotherapy. In: Lambert M.J. (Ed.), Bergin and Garfield's handbook of psychotherapy and behavior change. New York: Wiley & Sons.

Claus, R. E., Kindleberger, L. R., & Dugan, M. C. (2002). Predictors of Attrition in a longitudinal study of substance abusers. *Journal of Psychoactive Drugs*, *34* (1), 69–74.

Cournoyer, L. G., Brochu, S., Landry, M., & Bergeron, J. (2007). Therapeutic alliance, patient behavior and drop-out in a drug rehabilitation programme: the moderating effect of clinical subpopulations. *Addiction*, *102*, 1960-1970.

Crits-Cristoph, P. & Connolly, M.B. (2003). Research developments on the alliance in psychodynamic psychotherapy. *Psychoanalytic Inquiry*, 23,332-349.

Darke, S., Campbell, G., & Popple, G. (2011). Retention, early dropout and treatment completion among therapeutic community admissions, *Drug and Alcohol Review*. DOI:10.1111/j.1465-3362.2011.00298.x

Dawson, D. A., Grant, B. F., Stinson, F. S., and Chou, P. S. (2005). Psychopathology associated with drinking and alcohol use disorders in the college and general adult populations. *Drug Alcohol Depend*. 77, 139–150

EMCDDA (2008). Towards a better understanding of drug-related public expenditure in Europe, OPOCE, Luxembourg.

Fiorentine, R., Anglin, M. D., Gil-Rivas, R., & Taylor, E. (1997). Drug treatment: explaining the gender paradox. *Substance Use and Misuse*, 32, 653–678.

Fiorentine, R., Nakashima, J., & Anglin, M. D. (1999). Client engagement in drug treatment. *Journal of Substance Abuse Treatment*, *17*, 199–206.

Flynn, P. M. & Brown, B. S.(2008). Co-occurring disorders in substance abuse treatment: Issues and prospects. *Journal of Substance Abuse Treatment*.

Herbeck, D.M., Fitek, D.J., Svikis, D.S., Montoya, I.D., Marcus, S.C., & West, J.C. (2005). Treatment compliance in patients with comorbid psychiatric and substance use disorders. Am J Addict, 14 (3): 195-207

Hser, Y., Grella, C.E., Hubbard, R.L., Hsieh, S.C., Fletcher, B., Brown, B.S., et al. (2001). An evaluation of drug treatments for adolescents in U.S. cities. *Archives of General Psychiatry*, 58, 689-695.

Jackim, L.W. (2005). Entering the diagnostic debate. *Behavioral Healthcare Tomorrow*, 12-16.

Jentsch, J., & Taylor, J. R. (1999). Impulsivity resulting from frontostriatal dysfunction in drug abuse: Implications for the control of behaviour by reward-related stimuli. *Psychopharmacology*, 146(4), 373-390.

Jones, A., Donmall, M., Millar, T., Moody, A., Weston, S., Anderson, T., Gittins, M., Abeywardana, V., & D'Souza, J. (2009). The Drug Treatment Outcomes Research Study (DTORS): Final outcomes report. *Research Report 24*.

Kreyenbuhl, J., Nossel, I. R., & Dixon, L. B. (2009). Disengagement from mental health treatment among individuals with schizophrenia and strategies for facilitating connections to care: A review of the literature. *Schizophrenia Bulletin*, 35(4), 696–703.

Krueger, R. F., & Eaton, N. R. (2010). Personality Traits and the Classification of Mental Disorders: Toward a More Complete Integration in DSM–5 and an Empirical Model of Psychopathology. *Personality Disorders: Theory, Research, and Treatment*, Vol.1, *No.2*, 97–118.

Krueger R F, Markon K E, Patrick C J, Benning S D, Kramer M D (2007). Linking antisocial behavior, substance use, and personality: an integrative quantitative model of the adult externalizing spectrum. J *Abnorm Psychol* 116:645–666.

Livesley, W.J. (2007). A framework of integrating dimensional and categorical classification of personality disorders. *Journal of Personality Disorders*, 21 199-224.

Livesley, W. J., & Jang, K. L. (2000). Toward an empirically based classification of personality disorder. *Journal of Personality Disorders*, 14, 137–151.

Lowe, J.R., & Widiger, T.A. (2009). Clinicians' judgments of clinical utility: A comparison of the DSM-IV with dimensional models of general personality. *Journal of Personality Disorders*, 23, 211-229.

McLellan, A.T., Luborsky, L., & Woody, G.E. (1983). Predicting response to alcohol and drug abuse treatments: role of psychiatric severity. *Arch Gen Psychiatry*, 40:620-625.

Meier, P. S. & Barrowclough, C. (2009). Mental health problems - are they or are they not a risk factor for dropout from drug treatment? A critical review of the evidence. *Drugs Prevention and Policy*, *16*, 7-38.

Mulder, R. T., Frampton, C. M. A., Peka, H., Hampton, G., Marsters, T. (2009). Predictors of 3-month retention in a drug treatment therapeutic community. *Drug Alcohol Rev*, 28:366–71

Muller, S.E., Weijers, H.G., Boning, J., & Wiesbeck, G.A. (2008). Personality traits predict outcome in alcohol dependent patients. *Neuropsychobiology*, 57, 159-164.

O'Brien, A., Fahmy, R., & Singh, S.P. (2009). Disengagement from mental health services. *Social Psychiatry and Psychiatric Epidemiology*, 44,558–568.

Samuel, D. B., Simms, L. J., Clark, L. A., Livesley, W. J., & Widiger, T. A. (2010). An Item Response Theory Integration of Normal and Abnormal Personality Scales. *Personality Disorders: Theory, Research, and Treatment*, Vol.1, *No.1*, 5–21.

Schulte, S. J., Meier, P. S., Stirling, J., & Berry, M. (2010). Unrecognised dual diagnosis: A risk factor for dropout of addiction treatment, *Mental Health and Substance Use: dual diagnosis*, Vol 3, No. 2, 94-109.

Simpson, D. D. (2004). A conceptual framework for drug treatment process and outcomes. *Journal of Substance Abuse Treatment*, 27, 99-121.

Simpson, D. D. (2001; 2005). TCU Treatment Assessment Forms. Fort Worth: Texas Christian University, Institute of Behavioral Research.

Simpson, D. D., & Dansereau, D. F. (2008). Assessing organizational functioning as a step toward innovation. *NIDA Science & Practice Perspectives*.

Simpson, D. D., Joe. G. W., Brown, B. S. (1997). Treatment retention and follow-up outcomes in the Drug Abuse Treatment Outcome Study (DATOS). *Psychology of Addictive Behaviors*, *11*:294–307.

Simpson, D.D., Joe, G., & Broome, K. (2002). A national 5-year follow-up of treatment outcomes for cocaine dependence. *Archives of General Psychiatry*, 59(6), 548-554.

Simpson, D. D., Joe, G. W., Rowan-Szal, G. A., & Greener, J. M. (1997). Drug abuse treatment process components that improve retention. *Journal of Substance Abuse Treatment*, 14, 565-572.

Stark, M. J. (1992). Dropping Out of Substance Abuse Treatment: A Clinically Oriented Review. Clinical Psychology Review 12 (1): 93-116.

Sweet, C.,& Noones, J.(1989).Factors associated with premature termination from outpatient treatment. *Hospital and Community Psychiatry*, 40, 47–51.

Tsaousis, I. (2002). The traits personality questionnaire (TPQue). In B. De Raad & M. Perugini (Eds.), *Big five assessment* (pp.237–260). Göttingen: Hogrefe & Huber.

Van den Bosch, L.M.C. & Verheul, R. (2007) Patients with Addiction and Personality Disorder: Treatment Outcomes and Clinical Implications. *Curr Opin Psychiatry*, 20(1):67-71.

Verheul, R. (2005). Clinical utility of dimensional models for personality pathology. *Journal of Personality Disorders, 19,* 283-302.

Verheul, R., Andrea, H., Berghout, C.C., Dolan, C., Busschbach, J. V., van der Kroft, P. J. A., Bateman, A. W., & Fonagy, P. (2008). Severity Indices of Personality Problems (SIPP-118): Development, Factor Structure, Reliability, and Validity. *Psychological Assessment*, Vol.20, *No.1*, 23–34.

Is attention influenced by bilingualism in young children of low socioeconomic status?

Aristea I. Ladas¹, Ana B. Vivas², and Michael Siegal³

¹ South East European Research Center (SEERC), University of Sheffield, arladas@seerc.org

² Department of Psychology, the International Faculty of the University of Sheffield City College, vivas@city.academic.gr

³ Department of Psychology, the University of Sheffield, m.siegal@sheffield.ac.uk

Abstract. As bilingualism is nowadays considered to be a universal characteristic (Siegal et al., 2010), the need to clarify its exact effects on cognition has increased. Although recent research, involving participants of middle or high socioeconomic status (SES), has claimed for a bilingual advantage in non-linguistic attention tasks (e.g. Bialystok & De Pape, 2009; Kharkhurin, 2010), our previous two experiments with bilingual adults and children of low SES have shown no effect of bilingualism on attention. This suggests that maybe low SES can attenuate this bilingual advantage. In the present experiment, we asked whether a bilingual effect may be detected in even vounger children, whose attentional capacities are still developing. Secondly, we assumed that maybe a bilingual effect may be evident in a task that taps on inhibition of a linguistic nature, the Scalar Implicatures task (SI). Participants were young children. They were screened for general intelligence and vocabulary proficiency. The computerized Attentional Networks Task adopted for children (ANT Child; Rueda et al., 2004) was used to assess the three main attention networks, alerting, orienting and executive control of attention. Unlike other studies, we employed a more objective measure of language skill; a version of the computerized numerical languageswitching task of Meuter & Allport (1999) adopted for children. Finally, the SI task of Papafragou and Musolino (2003) was used to measure language inhibition. Comparative analyses did not show any significant effect of bilingualism in attention, and language inhibition when controlling for SES. However, there was a significant relationship between executive attention and bilingual control of languages. These findings support our previous conclusion that the bilingual advantage is eliminated when SES is controlled for. Keywords: bilingualism, attention, SES, children, Scalar Implicatures

1. Introduction

Nowadays, bilingualism is a universal norm (Siegal et al., 2010). Consequently, the need to clarify the possible influences of bilingualism on cognition has increased. As opposed to earlier studies, claiming for a bilingual disadvantage in cognition (Ivanova & Costa, 2008), more recent evidence supports a bilingual benefit in non-linguistic cognitive tasks (e.g. Bialystok & De Pape, 2009; Kharkhurin, 2010; Poulin-Dubois, Blaye, Coutya & Bialystok, 2011), with bilinguals outperforming monolinguals on tests of cognitive control.

However, most of the studies conducted so far have not controlled well for nonlinguistic factors, such as socio-economic status (SES), that could provide an alternative interpretation of the bilingual advantage (Morton & Harper, 2007). Moreover, while executive attention (e.g. conflict resolution; goal-directed attention) has been relatively well studied in bilingualism, the two other functions of attention, alertness and visuospatial orienting, have been largely ignored. To our knowledge, only two studies have investigated the three main attention networks (i.e. executive attentional control, alertness and orienting; Posner & Petersen, 1990) in bilingual individuals (Costa, Hernández & Sebastian-Gallés, 2008; Hernández, et al., 2010).

Hence, in our previous two experiments, we investigated the possible effects of bilingualism in the three main attention functions, in young adults (exp. 1) and children aged 8 to 10 years old (exp. 2), all of low SES. Attention did not seem to be influenced by bilingualism in either of these two experiments, thus suggesting that low SES possibly attenuated the so-called "bilingual advantage" in attention.

In an effort to provide stronger evidence on the existence, or not, of a possible bilingual benefit on attention, a third experiment was conducted. In this study, we asked whether a bilingual effect may be detected in even younger children, whose attentional capacities are still developing. Very young age could render any effect of bilingualism on attention more salient. Secondly, we assumed that a bilingual effect may be evident in a task that taps on inhibition of a linguistic nature, the Scalar Implicatures task (SI). Additionally, we aimed at examining whether the amount of inhibition used in resolving conflicting information (measured by the Attentional Networks Task adopted for children; ANT Child; Rueda et al., 2004) or/and switching between pragmatic and semantic meaning of a sentence (measured by the SI task; Papafragou & Musolino, 2003) is related to the amount of inhibition a bilingual uses to effectively switch between two languages (measured by the numerical language-switching task of Meuter & Allport, 1999, adopted for children).

Hypotheses

1. The executive network of attention (ANT Child):

The bilingual benefit in performance may be more salient in children. Thus, bilingual children of low SES will resolve conflicting information faster than their monolingual peers of equally low SES.

2. The alerting network of attention (ANT Child):

Due to lack of previous developmental findings on bilingualism and the alerting function, and the strong influence of SES on the developing alerting system, we predicted that bilingual and monolingual children of low SES will show an alerting effect of a similar magnitude.

3. The orienting network of attention (ANT Child):

Due to lack of previous developmental findings on bilingualism and the orienting function, and the strong influence of SES on the developing orienting system, we assumed that no differences among bilinguals and monolinguals in the orienting effect will be found.

4. Inhibition of a linguistic nature (SI task)

According to previous studies using the SI task to measure the ability to switch between pragmatic and semantic meaning of a sentence of bilingual as opposed to monolingual children (Siegal eta l., 2010), bilingual children will provide more correct responses in the SI task than their monolingual peers.

2. Method

Participants

The bilingual group included 32 young children bilingual in Greek and Albanian, with low SES. The monolingual group was constituted of 32 young children, monolingual in Greek, who were matched on SES to the bilingual group. All participants were screened for general intelligence and vocabulary ability, with the Raven's Coloured Progressive Matrices (CPM; Raven, Court & Raven, 1986) for children and the Vocabulary subtest of the Weschler Intelligence Scale for Children-version III (WISC-III; 1991), in both Greek and Albanian for the bilingual group. for detailed information on the participants se Table 1. Written informed consent was obtained from all participants. Testing lasted approximately 40 minutes for the monolinguals and 50 minutes for the bilinguals.

Table 1. Summary of participants' ages, SES level, WISC Vocabulary and Raven's CPM raw scores in the monolingual (n=32) and bilingual groups (n=32) in Experiment 3.

	Monolingual M (SD)	Bilingual M (SD)	<i>t</i> -values
Age	6.41 (0.85) Min age 5 Max age 7.7	6.80 (0.68) Min age 5 Max age 7.8	t(62)= 2.022, p= .047
<u>SES</u>	3.48 (0.96) Min SES 2	2.42 (1.15) Min SES 0.5	<i>t</i> (62)= 4.006, <i>p</i> < .00001

	Max SES 5	Max SES 5	
<u>G-WISC</u> <u>Voc*</u> <u>A-WISC</u> Voc*	11.13 (4.94)	10.97 (5.33) 11.63 (4.01)	<i>t</i> (62)= .122, <i>p</i> > .05
Raven's CPM	20.13 (6.38)	19.41 (4.81)	<i>t</i> (62)= .509, <i>p</i> > .05

* G-WISC Voc. = raw scores on the Geek version of WISC Vocabulary subscale * A-WISC Voc. = raw scores on the Albanian version of WISC Vocabulary subscale SES cutoff scores: 1 to 7= low SES, 7.5 to12= middle SES, $\leq 12.5 = high SES$

Material and Procedure

Each participant was tested individually, in a quiet room. A short description of the study, by withholding the exact aims to avoid demand characteristics, was firstly given, followed by the informed consent. The demographics and language background questionnaire were then be filled in by the participant, with the experimenter present so that any misunderstood question can be explained. The intelligence measures followed (Raven's SPM; WAIS-III Vocabulary test).

Furthermore, the computerized tasks were executed (language-switching as the measure of level of bilingualism; the ANT task). For both tasks, instructions were given orally and in written. The order of task administration was counterbalanced for all participants.

Intelligence and Vocabulary measures

In bilingual studies, it is essential to match participants for IQ (e.g. Colzato et al., 2008). Importantly, a widely used measure of intelligence has not yet been standardized for the Albanian population (e.g. Wasserman et al., 2000; Zimmerman et al., 2006). Thus, following previous studies of the bilingual effect on attention processes, though in adults, (Colzato et al., 2008; Treccani, Argyri, Sorace & Della Salla, 2009), Raven's Coloured Progressive Matrices (CPM; Raven, Court & Raven, 1986) for children was used to measure general intelligence in adults (see Table 1).

Secondly, as an indicator of vocabulary proficiency, the Vocabulary subtest of the Weschler Intelligence Scale for Children-version III (WISC-III; 1991) was administered in both languages (i.e. Albanian and Greek). The WAIS has not been standardized for Albanians. To partially compensate for this limitation, a psychologist of Albanian nationality scored the bilingual participants on the Albanian vocabulary test, from their recorded answers.

Demographics and Language background questionnaire

This questionnaire was based on previous studies of bilingualism and its hypothesized effect on cognitive processes, which have demonstrated to reliably and validly measure these demographic and linguistic aspects (Abedi, Lord & Plummer, 1997; Brown, Bown & Eggett, 2009; Costa, Garrat & Kelly, 2008; Gullberg & Indefrey, 2003; Portocarrero et al., 2007).

Although the questions were simply translated in Greek from the English language questionnaires, the language in which a self-rated questionnaire on bilingual language background is written does not appear to have an effect on the ratings (Delgado et al., 1999).

Computerized tasks

Both the ANT and the Language-switching task were displayed on a 15 inches monitor of a laptop PC, using the E-Prime 1.1 (2002) software.

The ANT task

The ANT was adopted from Rueda et al. (2004). The target was a fish, pointing either left or right. The target was always presented centrally, either alone (neutral trial) or flanked by four identical fish, according to the condition (congruent or incongruent trial). There were 4 cue conditions (see Figure 1).

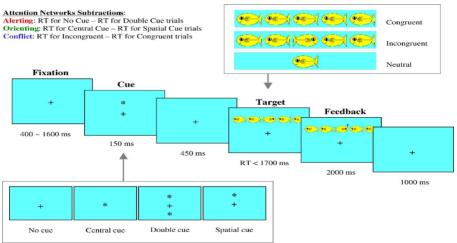


Fig. 1. All conditions and an example trial of the ANT Child. Adopted from Rueda et al., 2004.

One of the 12 following conditions was equally represented in each trial: 4 cueconditions (central cue, double cue, spatial cue, no cue) x 3 flanker conditions (congruent, incongruent, neutral). Response latencies (RT) and accuracy (errors) were recorded. For a depiction of a typical trial see Figure 2. Completion time was approximately 15min.

The Language-switching task

This task was similar to that of Meuter and Allport (1999), though adopted for children. The target stimuli were Arabic digits (1-9). The background was either a Greek or an Albanian colored flag (depending on the condition) serving as the language cue (i.e. prompted participants to read the digit in either language) see Figure 3). A microphone, connected to a voice key, was used to respond to the target. Participants were instructed to respond (i.e. read aloud the digit on the screen) in the language suggested by the language-cue (flag). They were encouraged to respond as quickly and accurately as possible.



Fig.3. Example of stimuli serving as language cues in the language-switching task.

The task was self-paced (i.e. each trial lasted until response to target) and the next trial onset took place 400ms after response (triggering of voice-key). Response latencies (RT) were recorded by the software. Digits were presented serially, from 1 to 9. All trials required the participant to switch from L1 (Albanian) to L2 (Greek) and vice-versa (i.e. switch trials). There were two experimental blocks. In the first block, the first language cue was the Albanian flag. After all 9 digits were presented serially in 9 switch trials, the language cue for digit "1" was the Greek flag. The second block was identical to the first one.

The Scalar Implicatures task

A version of Papafragou and Musolino (2003) SI task will be used, which has been shown to be appropriate for measuring pragmatic competency of Greek children, of 5 years of age. Three types of scales will be used: (a) a scale with numerical terms (i.e. *three, two*), (b) a scale with terms of quantity (i.e. *all, some*) and (c) a scale with initiation/finalization terms (i.e. *start, finish*). Thus, the children will be tested on how they interpret three types of scalar terms: *two, some, start*. These terms will be presented in the following sentences:

(1) Two of the horses jumped over the fence

(2) Some of the horses jumped over the fence

(3) The girl *started* making the puzzle.

However, the actual videotaped stories that are going to be presented will be truthfully described by the stronger term of each scale:

(4) Three of the horses jumped over the fence

(5) All of the horses jumped over the fence

(6) The girl *finished* making the puzzle.

For example, in the story for sentence (1), actually three horses will manage to jump over the fence. Thus sentence (1), although partially true, is pragmatically

infelicitous for describing that story, as it implies that only two horses jumped over the fence. Instead, sentence (4) should have been used. If a child correctly interprets this implicature then, she/he will respond that (1) was not a good answer for describing the corresponding story. The same logic underlies the other two statements. Thus, we expect that if a child is sensitive to the scalar implicatures used in statements (1), (2), and (3), she/he will respond that all these statements are not a good way to describe the stories where (4), (5) and (6) are depicted.

A hand puppet ("Mickey"), manipulated by the experimenter, will be introduced to the child. The training phase will be as follows: The puppet will be presented with a toy-horse. The experimenter will then ask Mickey "What is this Mickey?" and the puppet Mickey will reply "It is a horse". Then, the experimenter will ask the child "Did Mickey reply well?". After that, the puppet will be shown a log and when asked, it will reply "It is a table". If the child cannot provide a correct response when asked, the experimenter will say "Mickey did not say that very well. This is a log". These training scenarios will be used to ensure that children will critically consider the puppet's statements to decide whether what he says is true or not, instead of assuming that everything Mickey says is only true or false.

After that, the testing phase will begin. Children will be presented with video-taped scenarios, whereby the puppet Mickey will be watching toys completing a task (see Table 1). The procedure followed will be identical to the training phase, with the exceptions that the stories will be played on the video and that no feedback will be given on the child's responses for the test trials. The experimenter will say "Now, shall we watch some stories and see whether Mickey replies well?". The video will then be played, showing one of the stories described in Table 1. The video will then be stopped, and the experimenter will ask the puppet Mickey she will be manipulating what she thinks happened in that story. The puppet will answer using the terms described in Table 1 and the children will then be asked whether he replied well. Scores of correct responses will range from 0 (wrong reply) to 1 (correct reply) and will be manually recorded by the experimenter.

Scalar terms	Story showed on video	Puppet's statements	
	Three horses jumped	<u>Two</u> horses jumped over	
	over the log.	the log.	
	Three rabbits went in the	<u>Two</u> rabbits went in the	
	house.	house.	
{three, two}	Three dinosaurs ate	nouse.	
	trees.	<u>Two</u> dinosaurs ate trees	
	<u>Three</u> playmobils	<u>Two</u> playmobils bough	
	bought dogs.	dogs.	
	All horses jumped over	Some horses jumped	
	the log.	over the log.	
	All rabbits went in the	Some rabbits went in th	
(all some)	house.	house.	
$\{all, some\}$		Some dinosaurs ate	
	<u>All</u> dinosaurs ate trees.	trees.	
	<u>All</u> playmobils bought	Some playmobils bough	
	dogs.	dogs.	
	The tiger <u>finished</u>	The tiger started	
	painting the balloons.	painting the balloons.	
{start, finish}	The tiger <u>finished</u>	The tiger started putting	
	putting the cars into the	the cars into the bag.	
	bag.	the cars into the bag.	
	The little girl <u>finished</u>	The little girl started	
	making the puzzle.	making the puzzle.	
	The little girl finished	The little girl started	
	eating her food.	eating her food.	

Table 2. Test trials: Description of stories depicted in the video for the SI task and

 Puppet's statements that the children will be asked to judge.

As can be clearly seen from Table 1, all test trials will be designed to elicit a "No" response from the child. Thus, in order to balance out "Yes" and "No" responses, as well as to ensure that the child can correctly accept or reject the puppet's statements, especially when these statements are accurate descriptions of the story depicted (thus eliciting a "Yes" response), control trials will be also included in the task (see Table 2). Each control trial will be presented after each test trial (in the order described in Tables 2 and 3).

opet's statement	s.	
<u>Scalar</u> terms	Story depicted	Puppet's statements
	The tiger bought <u>two</u> of the balloons.	The tiger bought <u>two</u> balloons

Table 3. Control trials: Description of videotaped stories along with the analogous

	balloons.	balloons
	The strong man lifted two	The strong man lifted two
{three,	of the bags.	bags.
two}	Donald found two of the	Donald found two
	animals.	animals.
	Donald played with two of	Donald played with two
	the cars.	cars
	The tiger bought some of	The tiger bought some
	the balloons.	balloons
	The strong man lifted some	The strong man lifted
{ <i>all</i> ,	of the bags.	some bags.
some}	Donald found some of the	Donald found some
	animals.	animals.
	Donald played with some of	Donald played with some
	the cars.	cars.
	Donald started putting the	Donald started putting the
	pens into the pencil-case.	pens into the pencil-case.
	Donald started cleaning the	Donald started cleaning
{start,	table.	the table.
finish}	The little girl started	The little girl started
	painting the picture.	painting the picture.
	The little girl started	The little girl started
	drinking water.	drinking water.

3. Results

The following results are preliminary.

The ANT Child task

The latencies and error rates of 32 bilingual and 32 monolingual children were analyzed.

Error analysis

A 3 x 4 x 2 mixed ANOVA was used to analyze mean accuracy scores, with Flanker type (congruent, incongruent, neutral) and Cueing type (auditory alerting, no, central, spatial) as the within-subjects factors and Language group (monolingual and bilingual) as the between-subjects factor. There was a main effect of cueing, [F(3,186)= 4.088, p= .008, η^2 = .062]. According to Bonferroni post-hoc, children's rate of correct responses was significantly higher in the Auditory-Alerting -cueing condition (M=0.95, SD=0.07) than in the No-cueing condition (M=0.93, SD=0.07), p=.025. Additionally, there was a significant main effect of Flanker, [F(2, 124)= 17.538, p <.00001, η^2 = .221]. Bonferroni post-hoc showed that the rate of correct responses was significantly lower in the Incongruent (M= 0.91, SD= 0.11) compared to the Congruent (M= 0.96, SD= 0.06), p < .00001, and Neutral flanker conditions (M= 0.96, SD=0.05), p < .00001, thus indicating that the incongruent flanker led to less efficient performance in terms of errors. No other main effects or interactions reached statistical significance. Most importantly, error rates were similar and did not differ significantly across Language groups (monolinguals = 6.1%, bilinguals = 5.4%), which is in line with a previous study on bilingual adults using the ANT adult (Costa, Hernández & Sebastián-Gallés, 2006).

Response Times analysis

A 3 x 4 x 2 mixed ANOVA was used to analyze mean correct RTs, with Flanker (congruent, incongruent, neutral) and Cueing (auditory-alerting, no, central, spatial) as the within-subjects factors and Language group (monolingual and bilingual) as the between-subjects factor (see Table 2). A significant main effect of Flanker was indicated, [F(2, 124)= 130.398, p< .00001, η^2 = .678]. According to Bonferroni posthoc, RTs in the Incongruent condition (M= 900.05, SD= 142.06) were significantly slower than for the Congruent (M= 811.02, SD= 142.08) and Neutral (M= 780.67, SD= 136.28) conditions, p< .00001, irrespectively of Language groups. Thus, a Conflict effect was revealed for all participants. In addition, RT for Congruent trials were significantly slower than for the neutral trials, p< .00001.

There was also a main effect of Cueing, $[F(3, 186)=42.821, p<.00001, \eta^2=.409]$. According to Bonferroni post-hoc, performance in the Spatial-cueing condition was significantly faster than in all other cueing conditions. Specifically, RTs for the Spatial-cueing trials (M= 793.58, SD= 145.18) were significantly faster than for the Auditory-Alerting-cueing (M= 816.65, SD= 140.36), p=.041, and the No-cueing trials (M= 876.76, SD= 132.07), p<.00001. Importantly, RTs for the Spatial-cueing condition were significantly slower than for the Central-cueing condition (M= 835.33, SD= 154.72), p<.00001, thus indicating an Orienting effect across Language groups.

Additionally, RTs for the Auditory-Alerting-cue trials were significantly faster than for the No-cue trials, p < .00001, revealing the presence of an Alerting effect for

all participants. Similarly, performance was faster for the Auditory-Alerting-cue than for the Central-cue trials, p= .032. Finally, participants responded significantly faster in trials where they were presented with a central cue than in trials where they were not presented with any cue, p< .00001.

Flanker type	Language group	Warning type			
		No cue	Central	Double	Spatial
			сие	сие	сие
Congruent	Monolinguals	868.26	830.05	806.97	770.43
		(158.60)	(164.02)	(165.43)	(193.86)
	Bilinguals	862.84	788.56	794.03	766.98
		(134.86)	(135.00)	(138.99)	(156.30)
Incongruent	Monolinguals	937.46	908.48	909.64	878.03
		(144.13)	(156.85)	(162.53)	(159.57)
	Bilinguals	922.89	904.07	884.62	855.22
		(133.29)	(159.45)	(145.36)	(154.02)
Neutral	Monolinguals	845.36	795.73	746.83	766.52
		(156.18)	(164.01)	(151.35)	(167.74)
	Bilinguals	823.77	785.09	757.81	724.28
		(120.85)	(141.02)	(125.70)	(136.62)

Table 4. Means and (Standard Deviations) of RT in milliseconds for each Language group, in all flanker and Cueing conditions.

The analysis also indicated a significant interaction between Cueing and Flanker conditions, though this interaction has not been further analyzed yet.

Detailed assessment of the 3 attentional networks

According to Fan et al. (2002) and Rueda et al. (2004), the Conflict effect, reflecting the executive attention function, was calculated by subtracting mean RT in congruent trials from mean RT in incongruent trials. To obtain the Alerting effect, reflecting the alerting attention function, mean RT in auditory-alerting-cue trials were subtracted from mean RT in no-cue trials. Finally, to calculate the Orienting effect, reflecting the orienting network of attention, we subtracted mean RT in spatial-cue trials (i.e. trials were the cue was either above or below fixation) from mean RT in central-cue trials (see Table 3).

Executive network			
	Monolinguals (N= 32)	Bilinguals (N= 32)	
Trial Condition	<u>M (SD)</u>	<u>M (SD)</u>	
Incongruent	908.40 (145.96)	891.70 (139.87)	
Congruent	818.93 (153.45)	803.10 (131.71)	
Conflict Effect	89.47* (71.05)	88.6* (62.78)	
Alerting network			
	Monolinguals (N= 32)	Bilinguals (N= 32)	
Trial Condition	<u>M (SD)</u>	<u>M (SD)</u>	
No Cue	883.68 (143.74)	869.83 (121.19)	
Auditory-Alerting Cue	821.15 (154.84)	812.15 (126.56)	
Alerting Effect	62.53* (60.15)	57.68* (53.81)	
Orienting network			
	Monolinguals (N= 32)	Bilinguals (N= 32)	
Trial Condition	<u>M (SD)</u>	<u>M (SD)</u>	
Central Cue	844.75 (154.04)	825.91 (136.28)	
Spatial Cue	804.99 (152.37)	782.16 (139.09)	
Orienting Effect	39.76* (74.46)	43.75* (51.88)	

Table 5. Summary of performance of both Language groups in the ANT task (mean correct RT measured in milliseconds). The table is divided by the different conditions required to calculate performance in each main attention network.

(a) The executive function of attention (Conflict effect)

For the Conflict effect, a 2x3 mixed ANOVA was conducted, with Flanker type (Incongruent and Congruent trials) as the within-subjects factor and Language group (monolinguals and bilinguals) as the between-subjects factor. As expected by the earlier results, the analysis yielded a significant main effect of Flanker type [F(1, 62)= 112.873, p<.00001, η^2 = .645], with RT for Incongruent trials (M= 900.05, SD= 142.06) being slower than for Congruent trials (M= 811.02, SD= 142.08). Thus, a conflict effect was indicated for all participants, irrespectively of Language group. No other main effects and no interaction reached statistical significance. Interestingly, bilingual and monolingual children showed an almost identical Conflict effect (see Figure 1).

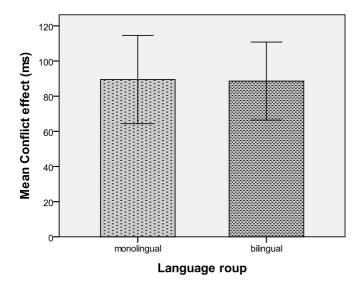


Fig.1. The Conflict effect was identical between monolingual and bilingual children. Bars represent SE.

(b) The alerting function of attention

Data were collapsed for Flanker and the results of 32 monolinguals and 32 bilinguals were submitted to a 2 x 2 ANOVA, with Cueing (No-cue and Auditory-Alerting-cue trials) as the within-subjects factor and Language group (monolinguals and bilinguals) as the between-subjects factor. Analysis of mean RT showed a significant main effect of Cueing [F(1, 62)=71.011, p < .00001, $\eta^{2}= .534$], with participants responding faster in the Auditory-Alerting-cue trials (M= 816.65, SD= 140.36) than in the No-cue trials (M= 876.76, SD= 132.07). Thus, an alerting effect was indicated as participants, irrespectively of Language group affiliation, were facilitated by an alerting cue. The interaction of interest though, among Cueing and Language group conditions did not reach statistical significance. Thus, it seems that Alerting was not modulated by bilingualism (see Figure 2).

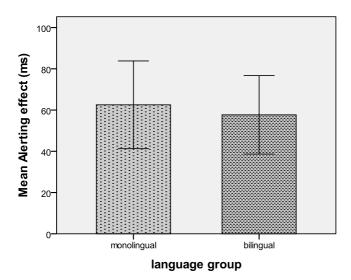


Fig.2. The mean Alerting effect (RT for No-cue – RT for Auditory-Alerting-cue trials) was similar in both bilinguals and monolinguals. Bars represent SE.

(c) The orienting function of attention

Data were collapsed for Flanker type and analyzed using a 2x2 mixed ANOVA, with Cueing (Central and Spatial) as the within-subjects factor and Language group (monolinguals and bilinguals) as the between-subjects factor. As reported earlier (see *General analysis*) a significant main effect of Cueing was revealed [F(1, 62)= 27.098, p<.00001, $\eta^2=.304$], with responses in trials with a Central, non-informative cue, being slower (M= 835.33, SD= 154.72) than responses in trials with a Spatial cue (M= 793.58, SD= 145.18), thus indicating an orienting effect regardless of Language groups. Neither the main effect of Language group nor the interaction among Cueing and Language group were significant, thus suggesting that bilingualism did not modulate the orienting attention function, as the orienting effect was similar for both bilinguals and monolinguals.

Language-switching task

To test the level of bilingual skill in the present bilingual group, mean RT in the Language-switching task were analyzed by paired t-tests, comparing response latencies to switch to L1 (Albanian) with response latencies to switch to L2 (Greek). Data from 32 bilingual children were included in the analysis. Results showed that mean RT to switch back to L1 (M= 2030.83, SD= 1045.28) were significantly slower than those required to switch to L2 (M= 1503.48, SD= 692.53), t(30)= 3.062, p=.005. Thus, an asymmetrical switch-cost was revealed from the analyses, indicating interference from L2 in switching to L1.

Furthermore, we subtracted mean RT to switch to L2 from mean RT to switch to L1, which provided us with an indicator of the amount of inhibition involved in such language-switching. Out of 32 bilingual children, 24 had Albanian as L1 and 8 had Greek as L1. We then entered this Language Inhibition variable to a correlation

analysis along with the Conflict, Alerting and Orienting effects provided by the ANT Child task. We expected that Language Inhibition would positively correlate with the Conflict effect. Additionally, given the mutual influence between the executive control of attention and the other two main attention networks, alerting and orienting (Callejas, Lupiáñez & Tudela, 2004; Callejas, Lupiáñez, Funes & Tudela, 2005), we speculated that maybe Language Inhibition will also be related to any or all of these attention processes.

	Conflict	Alerting	Orienting	correctSI
Language Inhibition	.409*	128	160	.014
Conflict		155	056	.213
Alerting			.065	186
Orienting				.165

Table 4. Correlations between the correct answers in the SI task (correctSI), the Conflict, Alerting and Orienting attention functions and Language Inhibition.

4. Discussion

Regarding the executive attention network, our results contradicted previous studies, by showing no difference in the performance of bilingual and monolingual children in the Conflict effect. This is in line with our previous findings in experiments 1 and 2. As for the other two main attention functions, alerting and orienting, these results confirmed our predictions, with the magnitude of the Alerting and the Orienting effects being roughly equal between monolingual an bilingual children. In sum, these results, in combination to those from our first and second experiments with adults and older children respectively, provide strong evidence on the role of SES in the so-called bilingual benefit in attention. That is, it seems that low SES attenuates the effect of bilingualism in attention found by previous studies.

Regarding the Language-switching task, firstly it seems that the bilingual children included in this study were dominant bilinguals, as a significant asymmetrical switchcost was found when switching from one language to the other. This opposes the parental statements on the Language background questionnaire, stating that these children were more or less balanced bilinguals and thus stresses the validity of the Language-switching task as an objective measure of level of bilingualism. Most importantly, Language Inhibition was positively related to the amount of conflict used to disentangle conflicting spatial information; this may suggest that maybe the mechanisms used for switching between languages and resolving conflicting information are indeed overlapping or even the same. Further analysis of these data is pending.

References

- Abedi, J., Lord, C., & Plummer, J. R. (1997). Final Report of Language Background as a Variable in NAEP Mathematics Performance. CSE Technical Report 429, Los Angeles: University of California.
- Bialystok, E. & DePape, A. M. (2009). Musical expertise, bilingualism and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 565-574.
- Brown, N. A., Bown, J. & Eggett, D. L. (2009). Making rapid gains in second Language Writing: A Case Study of a Third-Year Russian language course. *Foreign Language Annals*, 42, 424-452.
- Campbell, D. E. (2010). *Choosing Democracy: A Practical Guide to Multicultural Education (4th ed.).* UK: Pearson Education Inc.
- Colzato, L. S., Bajo, M. T., van den Wildenberg, W., & Paolieri, D. (2008). How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms. *Journal of Experimental Psychology: Learning, Memory* and Cognition, 34, 302-312.
- Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, 106, 59-86.
- Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language, 50,* 491-511.
- Fan, J., McCandliss, B. D., Sommer, T., Raz, M., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, 14, 340–347.
- Garratt, L. C., & Kelly, T. P. (2008). To what extent does bilingualism affect children's performance on the NEPSY? *Child Neuropsychology*, 14, 71-81.
- Gullberg, M., & Imdefrey, P. (2003). Language History Questionnaire. Retrieved November 2009, from http://www.mpi.nl/research/research-projects/the-dynamicsof-multilingual-processing/tools/Lang-Hist-Quest-Engl.pdf
- Hernández, M., Costa, A., Fuentes, L. J., Vivas, A. B., & Sebastián-Gallés, N. (2010). The impact of bilingualism on the executive control and orienting Networks of attention. *Bilingualism: Language and Cognition*, 13, 315-325.
- Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production. Acta Psychologica, 127, 277–288.
- Kharkhurin, A. V. (2010). Bilingual verbal and nonverbal creative behavior. International Journal of Bilingualism, 14, 211-226.
- Meuter, R. F. I. & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, 40, 25-40.
- Mezzacappa, E. (2004). Alerting, orienting, and executive attention: developmental properties and sociodemographic correlates in an epidemiological sample of young, urban children. *Child Development*, 75, 1373–1386.
- Morton, J. B. & Harper, S. N. (2007). What did Saymon say? Revisiting the bilingual advantage. *Developmental Science*, 10, 719-726.
- Papafragou, A. & Musolino, J. (2003). Scalar implicatures: experiments at the semantics– pragmatics interface. *Cognition*, 86, 253-282.
- Portocarrero, J. S., Burrightm, R. G., & Donovick, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology*, 22, 415-422.
- Posner, M. I., & Petersen, S. E. (1990). The attention systems of the human brain. Annual Review of Neuroscience, 13, 25–42.

- Poulin-Dubois, D., Blaye, A., Coutya, J., & Bialystok, E. (2011). The effects of bilingualism on toddlers' executive functioning. *Journal of Experimental Child Psychology*.
- Raven, J. C. (1958). The Standard Progressive Matrices. London: H. K. Lewis.
- Siegal, M., Surian, L., Matsuo, A., Geraci, A., Iozzi, L., Okumura, Y., & Itakura,S. (2010). Bilingualism accentuates children's conversational understanding. *Plos One*, 5.
- Treccani, B., Argyri, E., Sorace, A., & Sala, S. D. (2009). Spatial negative priming in bilingualism. *Psychonomic Bulletin & Review*, 16, 320-327.
- Weschler, D. (1997). *WAIS-III administration and scoring manual*. The Psychological Corporation, San Antonio, TX.
- Zimmermann, M. B., Connolly, K., Bozo, M., Bridson, J., Rohner, F. & Grimci, L. (2006). Iodine supplementation improves cognition in iodine-deficient schoolchildren in Albania: a randomized, controlled, double-blind study. *American Journal of Clinical Nutrition*, 83, 108-114.

Risky decision- making of Slovak undergraduate Students with and without migration experience.*

Eva Drobná¹, Viera Bačová¹

Department of Experimental Psychology, Slovak Academy of Sciences, Dúbravská cesta 9, 813 64 Bratislava, Slovakia edrobna@gmail.com

Abstract. The objective of this study is to investigate the interindividual and group differences in parameters of risky decisionmaking (value function, weighting function and loss aversion) according to Daniel Kahneman and Amos Tversky's Prospect Theory. Our sample consists of Slovak undergraduate students aged 23 - 27 with and without migration experience. We aim to find out whether there is a connexion between the level of willingness to risk and the migration tendencies among students. We use two experimental procedures - the replication of Tversky and Kahneman's PC procedure (used by authors to verify the postulates of Prospect Theory) and Sensation seeking scale. We hypothesize that students who will be more risk tolerant will be more willing to migrate abroad for study or work reasons and will also be a bigger sensation seekers.

Keywords: *Prospect Theory, Cumulative Prospect Theory, Parameters of risky decision- making, Students' migration*

1 Introduction

Man's decision- making. On the first look an ordinary and familiar daily process, but, on the other hand, it constitutes an inspiring subject of investigation for economical and psychological scientists, struggling to understand its patterns and relations for decades. From the economical point of view, decision- making had been considered a completely rational process for a long time. The normative economic theories (e.g. Expected Utility Theory) required people to act in accordance with the rules of logic. In spite of these expectations, experience and extensive evidence have shown that people do not act fully rationally in everyday life. For example, they do not evaluate probabilities objectively and rather overestimate the small ones and underestimate the high ones (Baláž, 2009). The inability to solve these issues paved

the way for the rise of new normatively- descriptive theory called the theory of prospect.

The main aim of our study is to find out, if students who are more risk tolerant will also be more willing to migrate abroad for study or work reasons. The literature about migration concentrates mainly on so called push (e.g. unemployment, poverty, war, population growth) and pull (e.g. better potential for employment, freedom, economic or political security, better quality of life) factors of migration – negative and positive elements which stimulate geographical movements of individuals and groups in relation to various socio- demographical indicators (www.iom.int). However, as King a Ruiz-Gelices (2003) suggest, standard academic migration literature does not pay much attention to students as migrants in practice. As there is in fact no experimental evidence investigating the specific relationship between risk perceiving and deciding to migrate by students, we concentrate in our study (Sec. 1.2) on common indicators which could clarify this examined relationship.

1.1 Prospect Theory and Cumulative Prospect Theory – psychological dimensions

Kahneman and Tversky present the Prospect Theory (1979) as a descriptive theory, since it replaced the rational, resp. logical rules of making decisions with the psychological ones, while keeping its mathematical exactness. In the Prospect Theory, the term "utility" has been replaced with "prospect". Utility is not the total wealth of a person but rather a change of assets upwards/downwards of the psychologically neutral reference point (usually the status quo). The term "prospect" is seen as a look out or hope, pinpointing the probability of the character of the result of the made decision which is, in addition, underlined by psychological factors (Baláž, 2009). According to Kahneman and Tversky (1979), people generally state the value of things in connection with the probability with which they will happen - they prefer smaller, but more certain gains to higher, but less certain ones. As the authors further declare, in choices between a number of risky prospects, we can observe several tendencies inconsistent with the basic principles of the Expected Utility Theory. Firstly, people undervalue the barely probable events compared with ones, which are definitive with certainty. This so called *certainty effect* contributes to risk aversion at choices meaning sure gains and to risk seeking at choices meaning sure losses. Next tendency is the *reflection effect* - in gains people are risk averse, while in losses, risk is a preference. This effect shows that risk aversion in the area of gains is accompanied with risk seeking in the area of losses - prospect reflection around zero swaps the order of preferences. The new element of PT compared with previous theories is a different attitude to probability. In PT probability is not understood as a number showing the objective/subjective probability. It is seen as a subjective decision weight (π), while incorporating psychological processes of decision making into it.

In 1992, Tversky and Kahneman published a new version of the Prospect Theory under the title of *Cumulative Prospect Theory*. The main difference to the original Prospect Theory is that it transforms the cumulative, not individual probabilities. The Cumulative Prospect Theory works with the same components of decision- making as

the initial Prospect Theory. It includes the weighting function and the value function, which is defined by gains and losses and is 2.25 steeper in the area of losses, compared to gains. To explain the characteristic curvatures of the value and the weighting functions, introduces the Cumulative Prospect Theory two principles. First it is the principle of *diminishing sensitivity*, which causes that the weighting function is concave close to the 0 and convex close to the 1. The second principle is the *loss aversion* - losses seem to be greater than the gains of the same value. In Cumulative Prospect Theory introduced Tversky and Kahneman a computer procedure, which aims at observing the parameters of risky decision- making (for detailed description see Sec. 2.1). These are especially important from the psychological point of view, since they give a statement on the particular specialties of each individual. If we want to state the prospect as decision- making criteria, it is necessary to know the span of these parameters in concrete person (Bačová, 2008). The parameters according to Tversky and Kahneman (1992) are following:

1. Parameter of the value function (Alpha, α) – in a concrete person, who makes a decision, his/her personal equation is created, between the relevant value and utility (stating, how far the considered individual is a "gamer" or a "coward").

2. Parameter of the weighting function (Gama, γ) – states how a concrete human assesses probability (how much his assessment differs from the objective probability). 3. Parameter Lambda, λ – states the rate of the deciding person's loss aversion.

As the results of Tversky and Kahneman experiment (1992) indicate, the most distinctive implication of the PT is the fourfold pattern of risky attitudes (see Table 1). Moreover, the results imply that all 25 participants in Tversky and Kahneman's study were risk averse in the case of positive prospects but in the case of negative prospects they were risk seeking.

	Small probabilities	Medium and large probabilities
Gains	Risk seeking	Risk aversion
Losses	Risk aversion	Risk seeking

Tab. 1: Risk attitudes depending up probability (Source: Kahneman & Tversky, 1979)

1.2 Risky decision- making in real-life situations: Are the migrating students bigger sensation seekers?

In this section of our study we concentrate on possible factors, which could influence the students' decision- making to migrate abroad for study or work reasons. First important factor is the risk perceiving. One of the sciences, which pay a lot attention to risk as one of the aspects influencing the decisions about migration, is economy. The main reason for it is that the decision to leave the home country temporary or as a permanency could influence the economic situation in the home country as well as in the target destination. The economic theory suggests that people always act rationally and they always take into account all costs including the risk by decision- making. Indeed, according to Williams & Baláž (2011), migration is always connected with some rate of uncertainty. Even a migrant who is provided with work, accommodation or social contacts at the target destination just before leaving the home country cannot be sure about his/her ability to adapt to the social and occupational conditions in the new environment. We could assume that if a risk averse person is not provided with sufficient information about the target destination he decides not to leave. This assumption was confirmed by Heitmueller (2005) who found out that the probability of migration decreases with the rate of risk aversion. In the context of our research conclusions of Donkers et al. (1999) and Dohmen et al. (2005) are also interesting. They suggest that the willingness to risk decreases with age. Teenagers and young adults are not only more risk tolerant but risk tolerance expresses itself also in their migration behaviour and intentions. According to Brezis & Soueri (2011), the ultimate factor influencing students to study abroad is the quality of education at the target destination. The study from Horvath and Zuckerman (1993) also serves as an inspiration for our research. The authors investigated relationships between sensation seeking and impulsivity, risk appraisal in several areas (crime, finance, social violations, sports and risk of AIDS from sexual activity) and risky behavior on the sample of undergraduate students. The authors have suggested that there are differences among people who strongly seek for sensation in risky activities (high sensation seekers) and people who rather avoid this kind of activities (low sensation seekers). Furthermore, sensation seeking and impulsivity predict risk appraisal and risky behavior. Indeed, they found out, that sensation seeking is positively correlated with risky behavior in all examined areas. We plan to apply the sensation seeking scale, used by the authors, in our research in order to examine whether our participants who will be more risk tolerant (according to replication of Tversky and Kahneman procedure, 1992) will also score higher on the sensation seeking scale.

2 Methods

Our research has the ambition to understand how people make their decisions in real- life situations. It is a part of research project of VEGA and SAS called *Interpretation of decision- making behavior: View of Economy and Psychology*. This project aims to contribute to the clarification of the nature and relationships between economic and psychological aspects of decision- making behavior. Keeping in mind the stated goals of this project and studied theoretical resources, we defined the following research questions and hypothesis in our study:

- ➢ Will the participants with migration experience differ from participants without migration experience in risky decision- making parameters?
- The participants with migration experience will be less risk averse than the participants without migration experience.
- More risk tolerant participants (according to replication of Tversky and Kahneman's experiment, 1992) will score higher on the sensation seeking scale than less risk tolerant participants.

2.1 Procedures

Sensation Seeking Scale

Sensation Seeking Scale (also called Impulsive Sensation-Seeking) contains 19 items in 2 subscales – impulsivity and sensation seeking.is and it is a part of Zuckerman-Kuhlman Personality Questionnaire (in complete wording available on the website http://grupsderecerca.uab.cat/). The scale serves for assigning the level of disposition to sensation seeking and is available on the website http://www.rta.nsw.gov.au.

Replication of Tversky & Kahneman's pc experiment (1992)

The main instrument of our research is the replication of Tversky and Kahneman's experiment (1992) in the form of a pc program, divided in two sections. In the first section, participant plays 28 games in 7 rounds. In some games he/she can win a fictive amount of money from 20 to 200ε , in others there is no win prize. The participant has to indicate what he/she prefers: either to play a risky game (e.g. 50% chance to win 20ε and 50% chance to win 0ε) or to pay off a guaranteed amount of money. And so the question is: at what amount of money do you prefer the payoff instead of playing the game? The instructions contain also the information about the value of each game. It is the amount of possible win prize (expressed in ε) multiplied by probability of this win prize (expressed on the continuum 0 to 1). Conclusions from the first section of the experiment will provide us with information which will help us calculate the parameters of value and weighting function for every participant.

The second section aims to find out the loss aversion rate for each participant and consists of 8 games, in which the participant can not only win fictive amount of money, but also can lose them. Games are based on the throwing of a coin. At throwing a coin, there are only 2 possible results. Each of them is 50% likely. The participant's task is to compare two games. Subsequently, he/she should state how high the sum of money would have to be to make the game equally attractive, regardless of the result. If the coin falls heads up, the participant loses $\in 0$, if it is tails, he wins $\in 0$. The participant neither wins, nor loses anything. This is the initial, neutral game. But, now we change the game – if it is heads, he/she loses $\in 12$. Such game is quite risky; hence a man would wish to receive some compensation for taking part in it. The question is: what would have to be the winning prize for you, so that the game is at least as attractive as the original and changed one and state what should be the winning prize of the changed game to make both games equally attractive.

2.2 Participants

Our sample consists of students aged 23 - 27 from the Faculty of Business Management at University of Economics in Bratislava. Participants (n = 120, 60 men and 60 women) will be divided in two groups: participants with migration experience and participants without migration experience. Participants with migration experience are students, who decided to leave their home country for 5 months/1semester or more in order to study (scholarship programme, language course) or work abroad (e.g.

during the summer holiday). Before leaving the home country they did establish their essentials (e.g. work, accommodation, meals etc.) in the terminal destination. Participants without migration experience are students who have never left their home country for 5 months/1semester or more in order to study or work abroad.

3 Conclusions

The aim of our study is to investigate the differences in parameters of risky decision- making according to the Prospect Theory. We anticipate that more risk tolerant students will be more willing to migrate abroad. Furthermore, we expect that more risk tolerant students will also be the greater sensation seekers.

* This Thesis Project was supported by the Scientific Grant Agency of Slovak Ministry of Education, Science, Research and Sport (VEGA) and Slovak Academy of Sciences (SAS) Grant Nr. 2/0121/10

References

Bačová, V. (2008) Classic normative theories of decision- making – psychological dimensions. In: Ruisel (Ed.), *Thinking – Personality – Wisdom* (pp. 77-96), Slovak Academic Press, Bratislava

Baláž, V. (2009) Risk and Uncertainty, Slovak Academy of Sciences Press, Bratislava

Brezis, E. S. & Soueri, A. (2011) Why do Students Migrate? Where do they Migrate to?, AlmaLaurea Working Papers no. 25, <u>http://www.almalaurea.it/universita/pubblicazioni/wp</u>, 24. 1. 2012

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J. & Wagner, G. G. (2005) Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validatet Survey, http://karlan.yale.edu/fieldexperiments/pdf/falk_risk%5B1%5D.pdf, 20. 3. 2011

Donkers, B., Melenberg, B. & Soest, A. (1999) Estimating Risk Attitudes Using
Lotteries:a LargeSampleApproach,http://www.springerlink.com/content/p7q1435km5716h70/, 16. 5. 2011

Heitmueller, A. (2005) Unemployment benefits, risk aversion, and migration incentives, *Journal of Population Economics*, 18: 93-112

Horvath, P. & Zuckerman, M. (1993) Sensation Seeking, Risk Appraisal, and Risky Behavior, *Person. Individ. Diff.*, Vol. 14, No. 1, pp. 41-52

Kahneman, D. & Tversky, A. (1979) Prospect Theory: An Analysis of Decision under Risk, *Econometrica*, Vol. 47, No. 2 (Mar., 1979), pp. 263-292.

King, R. & Ruiz-Gelices, E. (2003) International Student Migration and the European "Year abroad": Effects on European Identity and Subsequent Migration Behaviour, *International Journal of Population Geography*, Int. J. Popul. Georg. 9, pp. 229-252 (2003)

Tversky, A. & Kahneman, D. (1992) Advances in Prospect Theory: Cumulative Representation of Uncertainty, *Journal of Risk and Uncertainty*, 5:297-323

Williams, A. M. & Baláž, V. (2011) Migration, Risk, and Uncertainty: Theoretical Perspectives, <u>http://onlinelibrary.wiley.com/doi/10.1002/psp.663/pdf</u>, 19. 2. 2012

Sensation seeking scale, http://www.rta.nsw.gov.au/licensing/tests/driverqualificationtest/sensationseekingscal e/index.html, 4. 3. 2012 Zuckerman- Kuhlman Personality Questionnaire, http://grupsderecerca.uab.cat/zkpq/sites/grupsderecerca.uab.cat.zkpq/files/ZKPQ_AN <u>GLES%20web_0.pdf</u>, 15. 2. 2012

www.iom.int,

http://www.iom.int/jahia/Jahia/about-migration/key-migration-terms/lang/en#Pushpull-factors, 17. 6. 2012

Self-affirmation and doping behaviour

Vassilis Barkoukis¹, Lambros Lazuras², and Peter Harris³

 ¹ South Eastern Europe Research Centre (SEERC), vabarkoukis@seerc.org
 ² South Eastern Europe Research Centre (SEERC), llazuras@seerc.org
 ³ University of Sheffield (Psychology department), p.harris@sheffield.ac.uk

Abstract. Doping use hurts the image of sports, is against sport potentially irreversible health principles. and poses consequences to users. However, athletes act as non-users and they are not willing to participate in doping prevention or change strategies. Thus, in order for anti-doping campaigns and prevention strategies to be effective, it is vital to persuade athletes and exercisers to admit doping use, accept anti-doping campaigns and strategies, and participate with commitment and willingness. The application of self-affirmation manipulations that have been proven effective in other unhealthy behaviours, might offer an insight on the structure of future campaigns to combat doping and the way athletes and fitness exercisers should be approached in order to increase acceptance of the campaigns' message.

Keywords: self-affirmation, doping, elite sport, exercise

Introduction

Sport is universally promoted as the manifestation of health, excellence, hard work, fair play, and equality (Cooper, 2008). So far, doping prevention has largely focused on penalization of doping use and sanctioning of users. However, it appears that doping is still used by both professional athletes and exercisers (Striegel, Ulrich, & Simon, 2010), thus suggesting that a different doping use is illegal and severe punishment is involved, athletes and exercisers may be reluctant in expressing their actual intentions to use prohibited substances or admit their use to researchers and practitioners. Hence, they are not willing to participate in doping prevention or change strategies. Indeed, Petroczi et al. (2010) indicated that almost 50% of doping

users deny that they have used doping substances in self-report measures. Thus, in order for anti-doping campaigns and prevention strategies to be effective, it is vital to persuade athletes and exercisers to admit doping use, accept anti-doping campaigns and strategies, and participate with commitment and willingness.

1.1 Doping in sport

Doping use has been of outmost importance for policy-makers aiming to reduce unethical practice in sports, and to minimize the health risk imposed to athletes who use prohibited substances. Nevertheless, the recurring concern over doping has not been followed by systematic research on the psychosocial processes that underlie its onset and maintenance. Being conceptually similar to other forms of maladaptive behavior (e.g., cheating), doping can be adequately explained in terms of psychosocial processes like the formation of salient beliefs towards the use of prohibited substances; appraisal of behavioral control or efficacy over the use of prohibited substances; and the acquisition/internalization of social normative beliefs around the use of prohibited substances in sports (Lucidi, Zelli, Luca, Russo & Violani, 2008). Consideration of social norms relates both to subjective beliefs of acceptability or prevalence of doping, as well as perceptions of doping prevalence among athletes in general, and actual doping use at a team level (Lazuras, Barkoukis, Rodafinos & Tsorbatzoudis, 2010). While the former concept relates to a social cognitive process, the latter identifies the importance of social contexts wherein doping intentions and behavior are encouraged. Moreover, an athlete's overall inclination to engage in doping can be specified within broader motivational processes like motivational regulations, achievement goal and sportspersonship orientations (Barkoukis, Lazuras, Tsorbatzoudis & Rodafinos, 2011).

1.2 Self- affirmation

Self-affirmation theory (Steele, 1988) offers a sound theoretical basis for increasing the persuasive ability of a health-related message. According to this theory, individuals possess a self-image characterized by positive traits such as stability, free will and choice, kindness, a sense of being able to control personally significant outcomes, etc. A focal element of Steele's self-affirmation theory is 'self-integrity,' which reflects the experience of the self as 'adaptively and morally adequate.' According to Sherman and Cohen (2006), self-integrity comprises of five central domains: values (e.g., religion), social identities (e.g., being a good teammate in a football team), roles (e.g., being a loving wife), central beliefs (e.g., being prochoice), and goals (e.g., being healthy). When one of the aforementioned domains becomes threatened, then the person perceives their self-integrity to be in danger, and takes action in order to eliminate the threat. In this case, self-affirming restores an individual's self-integrity by reflecting upon one's cherished values, actions, or attributes.

Research evidence has indicated that manipulations based on self-affirmation effectively promote greater health risk information acceptance and reduce active message derogation in those at risk (Harris & Epton, 2009). Also, manipulations based on self-affirmation can increase reported negative affect in response to the

message. This was evident in numerous health-related behaviours such as smoking, HIV prevention, caffeine consumption, etc. Self-affirmation is effective in increasing message acceptance (promotes open- mindedness), reducing message derogation, increasing negative affect following health information, and increasing readiness to embrace risk-confirming or threatening material (Harris & Epton, 2009). According to Harris and Epton, evidence about how self-affirming promotes message acceptance and the role of significant others is still limited. Several mediators such as mood, state self-esteem, confidence and self-certainty have been examined, but findings across studies have not been consistent (Sherman & Cohen, 2006).

In addition, there is no thorough evidence on the mechanism through which self-affirmation manipulation influence the decision-making process that leads to behaviour (Harris & Epton, 2009). The theory of planned behaviour (TPB) has been repeatedly proven effective in describing a decision-making process across a range of health-related behaviours (Mullan & Wong, 2010). According to TPB, attitudes towards the behaviour, subjective norms, and perceived behavioural control form intentions towards the behaviour. In turn, intentions predict actual behaviour (Ajzen, 2002). TPB has provided a sound and comprehensive framework in understanding the mechanism leading to actual behaviour. Research evidence indicated that both intentions and their predictors are liable to change due to the use of self-affirmation manipulations. Intriguingly, although there is good evidence for positive changes in intentions, evidence on predictors is more mixed. This inconsistency could be ascribed to the relatively few studies that have examined each predictor to date (see Epton & Harris, 2009).

1.3 Self-affirmation and doping behaviour

So far, self-affirmation theory has been extensively applied in unhealthy behaviours. Yet, there is no evidence regarding illegal behaviours. Doping use is an illegal behaviour with also a potential harm to the athletes' health. European Union, World and National Anti-Doping Agencies and sport federations try to develop campaigns to combat doping as it threatens the spirit of sports and athletes' health. Yet, all these campaigns have failed and the number of athletes using doping substances increases. Also the use of such substances from fitness exercisers to improve physique emerged as a new trend. Clearly, the developed campaigns were not effective. This could be attributed to the fact that athletes and fitness exercisers are not susceptible to these campaigns, act as never doping users and resist changing. Thus, the application of self-affirmation manipulation that have been proven effective in other unhealthy behaviours, might offer an insight on the structure of future campaigns to combat doping and the way athletes and fitness exercisers should be approached in order to increase acceptance of the campaigns' message. Furthermore, so far self-affirmation theory has been applied to everyday life people in order to change an unhealthy habit. But there is no evidence on the effectiveness of the theory in changing a goal-directed behaviour, such as doping use. Moreover, change this behaviour to people who wish to benefit from it without taking into account the possible health hazards of misuse (i.e., athletes and fitness exercisers).

Thus, future research should investigate the effect of self-affirmation manipulations on the decision-making process (i.e., planned behaviour theory),

response biases (i.e., social desirability) and affect (i.e., regret) on elite athletes and fitness exercisers. Important research questions would involve:

a) Will self-affirmation manipulations influence the decision making process leading to doping use? More specifically, can self-affirmation manipulations create negative attitudes towards doping use, lower levels of perceived behavioural control and subjective norms and lower intentions to use doping substances in the future in both athletes and fitness exercisers,

b) Will self-affirmation manipulations decrease the social desirability scores in both athletes and fitness exercisers?

c) Will self-affirmation manipulations increase negative feelings, such as anticipatory regret, from doping use in both athletes and fitness exercisers?

References

- Ajzen, I. (2002). Residual effects of past on later behavior: Habituation and reasoned action perspectives. *Personality and Social Psychology Review*, *6*, 107-122.
- Cooper, C. E. (2008). The biochemistry of drugs and doping methods used to enhance aerobic sport performance. *Essays in Biochemistry*, 44, 63-83.
- Barkoukis, V., Lazuras, L., Tsorbatzoudis, H., & Rodafinos, A. (2011). Motivational and sportspersonship profiles of elite athletes in relation to doping behaviour. *Psychology of Sport & Exercise*, *12*, 205-212.
- Harris, P. R., & Epton, T. (2009). The impact of self-affirmation on health cognition, health behaviour and other health-related responses: A narrative review. *Social* and Personality Psychology Compass 3, 962–978.
- Lazuras, L., Barkoukis, V., Rodafinos, A., & Tsorbatzoudis, H. (2010). Predictors of doping intentions in elite-level athletes: A social cognition approach. *Journal of Sport & Exercise Psychology*, 32, 694-710.
- Lucidi, F., Zelli, A., Luca, M., Russo, P. M., & Violani, C. (2008). The socialcognitive mechanisms regulating adolescents' use of doping substances. *Journal* of Sports Sciences, 26, 447-456.
- Mullan, B., & Wong, C.L. (2010). Using the theory of planned behaviour to design a food hygiene intervention. *Food Control*, 21, 1524-1529.
- Petróczi A, Aidman EV, Hussain I, Deshmukh N, Nepusz T, et al (2010). Virtue or pretense? Looking behind self-declared innocence in doping. *PLoS One*, 5(5), e10457. doi:10.1371/journal.pone.0010457.
- Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defence: Selfaffirmation theory. Advances in Experimental Social Psychology, 38, 183–242.
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in Experimental Social Psychology*, *21*, 261–302.
- Striegel, H., Ulrich, R., & Simon, P. (2010). Randomized response estimates for doping and illicit drug use in elite athletes. *Drug and Alcohol Dependence*, 106, 230-232.

Danai Triantafylli1

¹ South-East European Research Centre (SEERC), Psychology Department, International Faculty, The University of Sheffield, datriantafyli@seerc.org

Abstract. Research findings indicate that drivers tend to overestimate their driving ability and safety. This documented optimism bias, in addition to negatively impacting engagement in precautionary driving behaviour, also interferes with the effectiveness of road safety campaigns. Studies regarding such cognitive biases have revealed that self-affirmation manipulations appear to be effective in reducing them and increasing message acceptance. The present study will be the first to investigate whether a value-affirmation manipulation could prove effective in reducing drivers' optimism bias and in increasing road safety message acceptance. Additionally, the contribution of a number of factors that could influence the impact of self-affirmation on actual driving behaviour will be investigated.

Keywords: Self-affirmation, drivers, optimism bias

1 Introduction

Road accidents are a major cause of death and injury universally, especially for those drivers aged under 25 (Clarke, Ward, & Truman, 2005). In the European Union (EU) during 2003 traffic accidents caused the death of roughly 47,000 people and the injury of 1.8 million people (Safetynet, 2005). According to numerous studies, at least half of road accidents are a result of reckless and risk-taking driving behaviours and could have been avoided given a more sensible driving. For example, Clarke and others (2005) report that 50% of accidents involving drivers under 25 years of age, are caused by excessive speed, alcohol consumption, and negligent driving. In Greece 29% of drivers involved in both fatal and not fatal accidents between 2001 and 2004 were found positive for alcohol consumption (Papadodima et al., 2008), while it is estimated that in the EU speeding is the most fundamental factor in 30% of all fatal accidents (European Road Safety Observatory, n.d., as cited in Forward, 2010). Moreover, in the past decade, instances of negligent driving have been multiplied due to the introduction of new electronic devices, often designed to be interactive and engaging (e.g. smart phones), in the car. The use of mobile phones alone, according to relevant studies, increases the likelihood of a crash by over 30% (Virginia Tech Transportation Institute, 2009, as cited in Wilson & Stimpson, 2010). Furthermore it is reported that, in the US, from 2001 to 2008 approximately 16,000

people have been killed due to drivers being distracted by the use of their mobile phones (for talking or texting; Wilson & Stimpson, 2010).

1.1 Drivers' optimism bias

However, the decisions made by the drivers to drink-drive, speed, or use their mobile phones while driving cannot be attributed to lack of knowledge about the possible consequences of those decisions; rather, it appears that drivers are aware that speeding increases the likelihood for a crash (e.g. Guerin, 1994), that mobile phone use while driving is risky (White, Eiser, & Harris, 2004), and also that they may belong to a high-risk age group (young drivers – Harré, Foster, & O'Neil, 2005). Moreover, while a considerable number of road safety campaigns informing the public about the consequences of various risky driving behaviours are being implemented every year worldwide, a significant proportion of drivers still engage in those risky behaviours. Seemingly, thus, increasing awareness and providing factual knowledge about risky driving does not seem to hinder or prevent risky driving patterns. This trend can be explained by the well-documented findings showing that drivers tend to underestimate their chances of being involved in a crash (e,g, Gosselin, Gagnon, Stinchcombe, & Joanisse, 2010), overestimate their driving abilities (Harré & Sibley, 2007), believe that they drive slower than average (e.g. Walton & Bethurst, 1998), and comply with driving regulations more than the average driver (e.g. Delhomme, 1991). Such beliefs are due to biased cognitive processing that people engage in, which in this case is optimism bias (Weinstein, 1980).

The engagement in such biased processing can also provide an account for the ineffectiveness of road safety campaigns, which is that "in order for a message to be received it must first overcome the cognitive distortions of the drivers" (Walton & McKeown, 2001, p. 639). In fact, the link between biased information processing and the ineffectiveness of road safety messages has also been empirically documented. Walton and McKeown (2001), reported that drivers who claimed that the average driver drives faster (and thus less safe) than themselves, but where mistaken in their belief (according to self-reported average speed), indicated that a presented safety campaign was not addressed to themselves.

1.2. Self-affirmation theory and cognitive biases

Nonetheless, drivers are not the only group of people to engage in such cognitive biases when they know they are choosing to engage in a risky behaviour, nor are they the only group that disregards the messages of health campaigns; similar biases have been also reported, for example, for smokers (e.g. Weinstein, Marcus, & Moser, 2005) and for individuals who engage in unprotected sex (Boney-McCoy, & Gibbons, 1999). A popular explanation for this phenomenon comes from Steele's self-affirmation theory (1988). According to this theory, individuals strive to preserve a self-image that depicts them as are "adaptively and morally adequate" (Steele, 1988, p.262). As such, when they are confronted with information they regard as threatening to that self-image, people tend to defend themselves by engaging in the

aforementioned biased cognitive processing, which allows them, at the very least, to minimize the personal relevance of that information. However, defensive cognitive processing is not the only 'weapon' that people have in their arsenal to defend their self image. Another contention of self-affirmation theory suggests that the individual does not defend the various aspects of their self-image separately, but rather they tend to defend it as a whole. Following from that suggestion, Steele (1988) argued that if an individual's self-image is threatened at a particular domain (e.g. their goal to remain healthy), then the person can adapt to this threat by affirming another important domain of their self-image (e.g. affirm their identity as a good parent). Studies have indeed demonstrated that people, when threatened, engage in spontaneous self-affirmations to alleviate the threat. Taylor's (1983, as cited in Steele, 1983) findings, for example, showed that female breast cancer patients tended to alter aspects of their lives in order to affirm central personal values unrelated to the domain of health (e.g. taking up creative writing or quitting an unfulfilling job). By making these changes in their lives, the participants of the study endeavoured to balance the threat posed to their self-image by the cancer.

Drawing form this capacity for 'fluid compensation'. Steele (1988) suggested that if a person is allowed the opportunity to affirm an important aspect of their self-image, then threats to another aspect would be better tolerated and the need to engage in cognitive biases would be eliminated. Experimental findings regarding the utility of self-affirmation manipulations in increasing health message acceptance have, indeed, supported this argument. Harris and Napper (2005), for example, used self-affirmation methodology to increase the acceptance of a health message regarding a link between increased alcohol consumption and breast cancer. Their results revealed that self-affirmed participants engaged in less defensive processing, accepted the health message more, and reported intending to reduce their alcohol consumption more than controls. However, the intervention did not result in actual behavioural changes (i.e. reduction of alcohol consumption). Armitage and his colleagues, (Armitage, Harris, Napper, & Hepton, 2008) designed a study that explored the effects of self-affirmation in an intervention regarding smoking (a widely practiced unhealthy behaviour) among factory employees in the UK. The results showed that self-affirmation decreased biased processing of health-threatening messages, which in turn influenced intention to change smoking behaviour. Moreover, van Koningsbruggen and Das (2009) designed a study to investigate the effects of self-affirmation in relation to level of risk. They used information regarding type 2 diabetes as the health-threatening message. They reported that self-affirmed participants who were at most risk were most benefited by the self-affirmation manipulation, since they derogated the health message less, and reported greater intentions of taking a type 2 diabetes risk online test than controls. Most importantly, though, those intentions were actually translated into behaviour, by taking the test.

Although affecting actual behaviour through self-affirmation methodology has proved to be challenging, the findings regarding its effectiveness in reducing cognitive biases have been consistently encouraging. Thus, it is possible that selfaffirmation would be also effective in reducing drivers' self-enhancement bias and promote acceptance of road-safety messages. In fact there are some evidence coming from studies investigating the effectiveness of negatively and positively framed road safety message, that provide some initial support for the aforementioned suggestion. Specifically, Harré, Foster, and O'Neill (2005), utilised in their study three negative videos showing instances when people lost their lives due to drunk driving, which were accompanied by declarations such as 'if you drink then drive, you're a bloody idiot', and three positive ones showing instance when people had chosen not to drunk drive, accompanied by declarations such as 'If you drink then don't drive you're a bloody legend'. Their results revealed that the positive road messages resulted in a decrease in self-enhancement bias as compared to the negative ones. Nonetheless, caution is warranted in interpreting these results since the study lacked a control group and it is not possible to determine whether the negative message increased biases or the positive message decreased them. To address this issue, Sibley and Harré (2009) replicated the aforementioned study and added a control group. They found that the positive message resulted in a significant decrease of such biases as compared to both a negative message condition and a control condition.

By examining the findings of the studies that investigated the effects of positive and negative road messages on drivers' optimism bias under the light of selfaffirmation theory, it appears possible that the positive messages operated as selfaffirmations for the drivers. Since it is not improbable that many drivers have chosen at some instances in their lives not to use their car after drinking, it is possible that the positive videos and messages reminded them of those instances and affirmed their sense of behaving smart, resulting in a decrease of their self-enhancement bias.

2. Conclusions

In the case of driving safety, defensive processing can have particularly precarious implications, resulting in people not having an accurate perception of their personal risk level of having an accident, which can result in lower motivation towards exercising precautionary driving behaviour (Klein, 1997), and in higher probability of engaging in risky driving behaviours such as speeding, or running red traffic lights (Fernandes, Job, & Hatfield., 2007). As such, the need for developing effective road safety campaigns and interventions that have a real potential of minimising the instances of reckless driving behaviours becomes urgent. Although numerous studies have investigated the effects of self-affirmation manipulation on health risk behaviours, the utility of this methodology in the domain of road safety has not, thus far, been investigated.

The purpose of this research project is to investigate the effectiveness of selfaffirmation manipulations on the cognitive biases and distortions that drivers display, as well as to explore the potential contribution of self-affirmation methods in developing more effective road safety campaigns. Until now researchers that have conducted studies evaluating self-affirmation manipulations, have targeted behaviours such as smoking, consumption of alcohol and caffeine over the recommended limit, and unhealthy diet; behaviours that compromise health in a 'slow-and steady' manner. This project will be the first to target a behaviour that can lead to injury or even immediate death. Moreover, since findings from previous studies indicate that selfaffirmation manipulations are limited in their capacity to amend actual behaviour, this project will also attempt to identify a (limited) number of factors that may hinder or enhance self-affirmation effects on actual behaviour.

References

Armitage, C. J., Harris, P. R., Napper, L., & Hepton, G. (2008). Efficacy of a brief intervention to increase acceptance of health risk information among adult smokers with low socioeconomic status. *Psychology of Addictive Behaviors*, *22*, 88–95.

Boney-McCoy, S., Gibbons, F. X., & Gerrard, M. (1999). Self-esteem, compensatory self-enhancement, and the consideration of health risk. *Personality and Social Psychology Bulletin*, *25*, 954–965.

Clarke, D.D., Ward, P., & Truman, W. (2005). Voluntary risk taking and skill deficits in young driver accidents in the UK. *Accident Analysis and Prevention*, *37* (3), 523-529.

Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When beliefs yield to evidence: reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin, 26* (9), 1151-1164.

Correll, J., Spencer, S. J., & Zanna, M. P. (2004). An affirmed self and an open mind: Self-affirmation and sensitivity to argument strength. *Journal of Experimental Social Psychology*, *40*, 350–356.

DeJoy, D. M. (1992). An examination of gender differences in traffic accident risk perception. *Accident Analysis & Prevention*, 24 (3), 237-246.

Delhomme, P. (1991). Comparing one's driving with others': assessment of abilities and frequency of offences. Evidence for a superior conformity of self-bias? *Accident Analysis and Prevention, 23* (6), 493-508.

Epton, T., & Harris, P. R. (2009). The impact of self-affirmation on health cognition, health behaviour and other health-related responses: a narrative view. *Social and Personality Psychology Compass*, *3* (6), 962-978

Fernandes, R., Job, R.F.S., & Hatfield, J. (2007). A challenge to the assumed generalizability of prediction and countermeasure for risky driving: different factors predict different risky behaviors. *Journal of Safety Research, 38* (1), 59-70.

Forward, S. E. (2010). Intention to speed in a rural area: Reasoned but not reasonable. *Transportation Research Part F, 13,* 223-232.

Gosselin, D., Gagnon, S., Stinchcombe, A., & Joanisse, M. (2010). Comparative optimism among drivers: an intergenerational portrait. *Accident Analysis & Prevention*, 42 (2), 734 – 740.

Guerin, B. (1994). What do people think about the risks of driving? Implications for traffic safety intervetions. *Journal of Applied Social Psychology, 24* (11), 994 – 1021.

Harré N., Foster, S., & O'Neil, M. (2005). Self-enhancement, crash-risk optimism and the impact of safety advertisements on young drivers. *British Journal of Psychology*, *96* (2), 215 – 230.

Harré, N., & Sibley, C. G. (2007). Explicit and implicit self-enhancement biases in drivers and their relationship to driving violations and crash-risk optimism. *Accident Analysis and Prevention*, *39* (6), 1155–1161.

Harris, P. R., & Napper, L. (2005). Self-affirmation and the biased processing of health-risk information. *Personality and Social Psychology Bulletin, 31*, 1250–1263.

Klein, W. M.(1997). Objective standards are not enough: Affective, self-evaluative, and behavioral responses to social comparison information. *Journal of Personality and Social Psychology*, *72(4)*, 763-774.

Lipkus, I. M., McBride, C. M., Pollak, K.I., Lyna, & P., Bepler, G. (2004). Interpretation of genetic risk feedback among African American smokers with low socioeconomic status. *Health Psychology*, 23(2), 178 – 188.

Papadodima, S. A., Athanaselis, S. A., Stefanidou, M. E., Dona, A. A., Papoutsis, I., Maravelias, C. P., & Spiliopoulou, C. A. (2008). Driving under the influence: A 7-year survey (1998 – 2004). *Forensic Science International*, *174* (2-3), 157 – 160.

Redelmeier, D. A., & Tibshirani, R. J. (1997) Association between cellular-telephone calls and motor vehicle collisions *New England Journal of Medicine*, *336*, 453-458.

Reed, M. B., & Aspinwall, L. G. (1998). Self-affirmation reduces biased processing of health-risk information. *Motivation and Emotion*, 22(2), 99–132.

Safetynet.(2005).Annualstatisticalreport2005.<http://www.erso.eu/safetynet/fixed/wp1/2005/sn-1-3-asr2005.pdf>

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defence: Self-affirmation theory. *Advances in Experimental Social Psychology*, *38*, 183–242.

Sibley, C. G., & Harré, N., (2009). The impact of different styles of traffic safety advertisement on young drivers' explicit and implicit self-enhancement biases. *Transportation Research Part F: Traffic Psychology and Behaviour, 12 (2),* 159–167.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In Berkowitz L.(Ed.), *Advances in Experimental Social Psychology* (Vol 21), Academic Press, New York, 261-302.

Svenson, O. (1991). Are we all less risky and more skilful than our fellow drivers? *Acta Psychologica*, 47 (2), 143-148.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological Bulletin*, *103* (2), 193-210.

van Koningsbruggen, G. M., & Das, E. (2009). Don't derogate this message! Self-affirmation promotes online type 2 diabetes risk test taking. *Psychology and Health*, *24*, 635–649.

Walton, D. & Bathurst, J. (1998). An exploration of the perceptions of the average driver's speed compared to perceived driver safety and driving skill. *Accident Analysis and Prevention*, *30* (6), 821-830.

Walton, D., & McKewon, P.C. (2001). Drivers' biased perceptions of speed and safety messages. *Accident Analysis and Prevention*, 33 (5), 626-640.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, 39 (5), 806-820.

Weinstein, N. D., Marcus, S. E., Moser, R. P. (2005). Smokers' unrealistic optimism about their risk. *Tobacco Control, 14,* 55-59.

White, M. P., Eiser, J. R., & Harris, P.R. (2004). Risk perceptions of mobile phone use while driving. *Risk Analysis*, 24 (2), 323-334.

Wilson, F. A., & Stimpson, J. P. (2010). Trends in fatalities from distracted driving in the United States, 1999 to 2008. *American Journal of Public Health, 100 (1),* 2213-2219.

Optimisation of solar thermal panel use

J.P.R. Williams¹, Christopher Jones², and Stephen B.M. Beck³

 ¹ Student (Departments of Psychology and Mechanical Engineering), The University of Sheffield UK, J.R.P.Williams@sheffield.ac.uk
 ² Supervisor (Department of Psychology), The University of Sheffield, UK, C.R.Jones@sheffield.ac.uk
 ³ Supervisor (Department of Mechanical Engineering), The University of Sheffield, UK, S.Beck@sheffield.ac.uk

Abstract Technologies that are more energy efficient are being sought due to the pressures emanating from: (i) Global Climate Change; (ii) diminishing stocks of fossil fuels; (iii) a massively expanding global population; and (iv) high energy consumption closely linked to modern life. Ensuring that new technologies are adopted by a large proportion of the population is dependent on having a clear understanding of consumer behaviour. One energy efficient technology is solar thermal panels (STPs) which have the potential to supply a large proportion of domestic hot water demand. Focus groups were conducted with three groups of consumers to better understand the public perception of the technology and any socio-economic barriers that might hinder the adoption of this technology. The consumers that were involved in the focus groups were: those that have STPs; those that are interested in investing in STPs; and those that have never thought of investing.

There is generally little understanding about how STPs work across all three groups, although this is less of a concern for those that own STPs. There were concerns about the effectiveness and the maintenance of the panels but most participants felt that receiving good information from either the local or national government would dispel any fears. The participants were wary of information originating from companies as it was felt this would be biased to try and sell a product. People were less concerned about their appearance or the fact that water would be heated by the sun. The information gathered in these focus groups can be used to inform a wider study in Sheffield and Thessaloniki to better understand consumer opinions of this technology. From these studies a clearer understanding of local and nation policy, to take full advantage of this technology, can be made.

Keywords: Solar Thermal Panels, Technology Acceptance Model, Focus Groups

1 Introduction

1.1 Solar Thermal Panels

Solar Thermal Panels (STPs) are collectors that are attached to the roofs of homes with water running through them which is heated by the sun's rays which can then be used in the home for baths and showers (Figure 1). In the UK heating water in homes accounts for approximately 25% of energy used in the home (Allen, Hammond, Harajli, McManus, & Winnett, 2010) and so any significant improvements in the technology used to heat domestic water will have a similarly significant energy improvement. STPs provide this improvement in energy efficiency in the heating of water for domestic use.

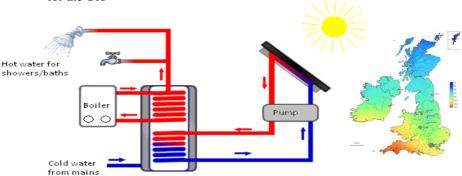


Figure 1: A solar thermal panel arrangement and a map of the solar radiation for the UK

At the start of 2011 there were approximately 100,000 STP systems installed on homes in the UK making it the most used form of micro-generating technology although this only accounted for 1TWh (DECC, 2011). With approximately 27 million homes in the UK this represents only 0.37% of homes in the UK and so there is clearly a large potential for the technology. To help ensure that STP technology is adopted widely a better understanding of the market in the UK is needed.

There are a number of existing models that describe consumer behaviour with respect to innovative technology (Technology Adoption Life Cycle, Diffusion of Ideas, Model of Adoption of Technology in Households) but the model that is discussed here is the Technology Acceptance Model (TAM) (Davis, 1989). The TAM has been applied to a number of technologies since it was devised by Davis in 1989 and is considered a robust model at describing actual behaviour of a technology but has yet to be applied to STPs. The TAM model is shown in Figure 2.

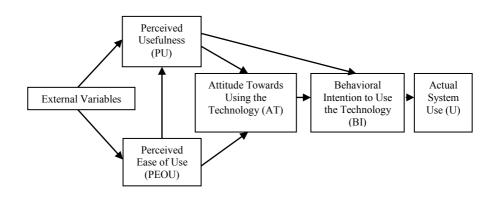


Figure 2: The Technology Acceptance Model (Davis, 1989)

Previous work concerning the TAM has been applied to a large number of technologies in the home (Porter & Donthu, 2006) and outside it (Legris, Ingham, & Collerette, 2003). The TAM has been studied in a large number of journals as detailed in (King & He, 2006). It has also been applied to different countries exploring any differences between the nations (Straub, Keil, & Brenner, 1997). However despite its widespread applications the TAM has never been applied to solar thermal panel technology. Previous work concerning domestic attitudes towards solar energy technology has tended to focus on solar photovoltaic technology (Faiers & Neame, 2006) although there has been scant research into consumer attitudes towards solar thermal technology (Woersdorfer & Kaus, 2011).

1.2 Aims of this study

This research was carried out with a few objectives in mind:

- Develop a better understanding of the drivers and barriers to consumer behaviour in the solar thermal panel market;
- Use the improved understanding to better inform a widespread study in the English city of Sheffield and the Greek city of Thessaloniki; and
- Incorporate the findings in to a novel solar thermal panel system.

The work presented here was performed to assess the main themes and opinions consumers had concerning energy used in a domestic setting. By assessing consumer Perceived Usefulness and Perceived Ease of Use a better understanding of consumer Attitudes Towards Using the Technology and their Behavioural intention to Use the Technology can be discerned.

1.3 Qualitatative Data

The use of focus groups to illicit public opinions is well established and has been used extensively in the energy sector (Itaoka, Saito, & Akai, 2011; West, Bailey, & Winter, 2010). A main benefit to this approach is that it allows a better understanding behind the 'why' people make the choices they do and how much emphasis is placed on those choices. Focus groups:

"...determine the perceptions, feelings, and thing of people about issues, products, services, or opportunities." (Krueger & Casey, 2009)

Focus groups are however limited in providing concrete analysis of generally held opinions if results are considered alone although focus groups can significantly direct further research.

1.4 Method

For this study, data came from focus groups conducted at the University of Sheffield after the participants had completed an online survey to distinguish which of the three categories they fell into:

- 1. Those who have solar thermal panels;
- 2. Those that do not have but are considering to buy solar thermal panels; and
- 3. Those that have never considered buying solar thermal panels.

The focus groups were held in groups of four or six with the discussion topics divided into five sections:

- 1. Introduction to energy;
- 2. Energy use discussion;
- 3. Displaying energy use;
- 4. National/global issues;
- 5. Solar thermal panels.

The focus groups each had a mix of genders, ages, ethnicities and backgrounds as well as residents with varying housing situations. The audio of the resulting conversations were audio recorded and transcribed for ease of analysis. The computer software NVivo 9 was used to analyse the transcripts of the conversations. The focus groups were led by a moderator who briefed the participants of the received ethics approval for the study and the reason for the focus group. The participants were also briefed about the nature of the present research although no specifics were given.

Results Introduction to energy

The discussion started with the participants talking about times when they have used a lot of energy in the home, which rooms in the home use a lot of energy, if they had noticed a change in the energy habits recently and if there were any concerns. The participants talked about specific occasions when they had used a lot of energy in the home and all indicated the living room and kitchen as being rooms in the home that predominantly use a lot of energy in the home, mainly due to cooking and the concentration of electrical equipment (such as TV's, stereos etc). This was also thought to be universal across all the groups with the general feeling that people in other countries use most of their energy in the home either in the kitchen or the living room. A few participants were foreign nationals and agreed that people from other countries would use energy in those two rooms as they would be using the energy for cooking or entertainment.

2.2. Energy usage discussion

Many participants talked of knowing they were energy inefficient with electrical items often left on standby rather than being turned off. Electrical items consuming electricity whilst being on standby was a common annoyance ('..especially when the reason for not turning off is because we know it will lose memory'. For the groups that heated their water using gas, leaving electrical equipment on was less of a concern for them, particularly energy efficient light bulbs. It was a common annoyance that the bulbs take a while to fully turn on and that was a reason why they were left on, however there were some dissenters to this viewpoint ('I know it's irritating at the beginning, but that I got used to').

The participants in all the groups were aware that they had recently reduced their energy usage due to a change in energy prices although this was less so in the group of people who had STPs installed on their home as they have a large proportion of their water heated from the sun and therefore haven't been affected as much by changes in energy prices. Energy prices were highlighted as the main reason that people have changed their energy consumption in the past ('I'm not sure that most people are conserving energy for the environment. I think it's more to do with money'. The groups also highlighted that there has been a change in energy consumption linked to both a change in technology and a change in attitudes. An improvement in technology has reduced energy consumption but so has a change in attitude to energy usage, particularly with children ('Sarah comes home and does tell me off for doing things, and then she forgets and does it herself!').

2.3. Displaying energy usage

Smart meters and displays that informed the participants of their energy usage had reduced some of their energy usage. This was felt especially when the cost of the energy usage was displayed. Smart meters were not known across all the groups however and there was some misinformation about them. However there were positive comments about their usage particularly as energy companies would be more accurate in their charges ('...it's going to put an end to estimated bills as well. They'll know exactly what you're using'). There were even some suggestions of an online website where a constant monitoring ('If there was an automatic system for logging all the meter readings on a regular basis directly onto my computer, that would save a lot of messing around').

People from all groups expressed a desire to have real-time feedback showing an instantaneous reading of how much energy is being consumed and a few thought it would change their energy behaviour ('Knowing how much you are using gives you an idea of how much you are spending, and curbs your energy usage' 'The washing machine and dryer use quite a lot. You knew this intellectually, but actually seeing a number attached is quite illuminating').

People on the whole do not pay much attention to the few places in the home that display their energy usage (electricity meters or gas meters). There was a strong agreement that they would be interested in knowing how much energy the individual devices in their homes consumed. People on the whole only noticed their energy consumption clearly when they paid their bills and although generally positive about smart meters there was a good deal of suspicion of energy companies and intentions of the energy companies installing these meters in their homes.

This suspicion was not reflected in the bills however, as the participants trusted the bills as being accurate of their energy used ('I would like to be able to actually check how much individual devices have taken. I think the bill is just the sum total, and it's very hard to figure out from the bill or the meter, really, where does that come from?'). In the group of participants who had STPs installed on their home they were aware of how much energy they consumed in the home (particularly their electricity usage) with some people sharing their energy usage with others outside their home ('I'm a member of a so-called Carbon Reduction Action... We keep data and compare each with another to see what the effects of particular houses, cars, and ways of doing things might be' 'I actually joined the iMeasure' 'I am a member of the Solar Farm Collective').

2.4. National/global issues

Participants in this section talked about their lack of knowledge of the government's policy towards global issues such as climate change and energy policy ('What the current government is actually doing I don't really understand'). There was agreement across all three groups that Climate Change was being exacerbated by anthropogenic activity. The participants talked about the responsibility of doing something about Climate Change resided with the individual and governments alike. Some participants raised the issue that governments have a problem with short-termism and not seeing beyond the next election so having a coherent policy and an education policy would be crucial to overcome this problem. A 'fear' of the electorate was said to be the reason why governments either did nothing to combat Climate Change or play down its concerns ('I've no confidence that they're doing anything serious').

With regard to other nations it was a general agreement that the UK government should be leading the way of smaller nations as well as larger nations ('If you're trying to persuade the Americans to use less energy the first thing to do is to show that you can do it yourself').

The government, it was felt should play a role in providing good advice to those wanting to invest in STPs and regulating STP companies ('It's like any kind of building work, it should be regulated').

2.5. Solar thermal panels

In the final discussion the first topic was to do with any concerns the participants might have with STPs. The main concerns were: their effectiveness; how much hot water could be produced; and if it would be hot enough to be used for baths and/or showers. These concerns were across all three groups and were shared even by the group of participants who currently own them. Some members of this group were still unsure as to how much hot water was generated by the panels or how they worked but the sentiment was treated light-heartedly by the group ('I think we need a 'solar power for dummies' book, actually!'). Other concerns were about the installation and maintenance of the panels ('In terms of it getting put up there. That'd have to be fairly quick and painless') although the group that has STP's installed reported a positive experience of STP companies ('Very good in our case. A personal relationship almost'). None of the groups were worried about water that was heated by the sun and most liked the idea ('Personally, I think it's a lovely idea).

The appearance was also mostly viewed favourably by Those That Have and Those That Are Thinking of categories with only a few concerns raised ('I like the idea. I don't like the visual impact.' The concerns centred on their appearance on older homes or in areas of particular beauty. There were more people in the Those That Are Thinking Category who found their appearance less favourable ('They are ugly looking' 'Bloody ugly') although this was a universal opinion for the group ('...solar panels, flat on the roof, not much uglier than anything else'). There were a few participants who felt that the appearance of the STPs would not affect their decision as to whether they would invest in the technology.

3. Discussion

The results garnered from the focus groups in this study appear to suggest that there is a link between a consumers perceived usefulness and perceived ease of use and their attitude towards the technology. As a minor study the work presented here can only be viewed in context but it does add weight to the argument that a questionnaire study would provide definitive answers to some of the issues raised. The focus groups also highlighted a good deal of misinformation and distrust of the government indicating that the improved education and availability of sound information to consumers would be advised. To that end further research is advised.

4. Conclusions

This paper analysed and commented on a number of focus groups of actors concerned with the solar thermal panel (STP) market in the UK. The focus groups involved three consumer groups; those that have STPs; those that are thinking of investing in STPs; and those that have never though t of investing in STPs. The focus groups showed that there is a lack of understanding of the technology with regard: it's operation; it's suitability; the required maintenance; and places where sound information can be received. Those that had never considered investing in STPs were less interested in the technology as a whole and would only install it if they could be guaranteed of its performance. Those that have STPs already installed on their homes were very positive about their experience and were keen to share this experience with others. They had a good experience of the STP companies this was probably due in part to their interest in the technology and their willingness to research the topic and inform themselves. They were less interested in features of the technology such as the appearance and the payback period. Money was a defining issue which would determine whether people would invest in the technology. There was little agreement as to whether STPs would increase or decrease the value of homes and for some that rented they simply did not know whether they could have the STPs installed.

5. Future Work

The work presented here has only started to fully understand the state of the solar thermal panel market in the UK. The qualitative data that has been analysed here can be expanded to encompass other important actors in the STP sector such as companies, politicians and community leaders. Once interviews are conducted with these actors a better understanding can be made of the drivers and potential stumbling blocks to ensuring that this technology is adopted widely. An extensive questionnaire could then be designed to include the work here and be distributed to a high number of households. It is intended that this study will be conducted in Sheffield, UK (population approximately 325,000) and Thessaloniki, Greece (population approximately 550,000). Repetition of the survey in the two cities will provide a clearer understanding of the solar thermal panel market in two cities that are comparable in size but are at different stages of development with the Greek market at least ten years more advanced than the UK market. Any lessons from the Greek market can then be learnt to inform and focus development of the UK market.

6. References

- Allen, S. R., Hammond, G. P., Harajli, H. A., McManus, M. C., & Winnett, A. B. (2010). Integrated appraisal of a Solar Hot Water system. [doi: 10.1016/j.energy.2009.11.018]. *Energy*, 35(3), 1351-1362.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.

DECC. (2011). UK Renewable Energy Roadmap. London.

- Faiers, A., & Neame, C. (2006). Consumer attitudes towards domestic solar power systems. [doi: 10.1016/j.enpol.2005.01.001]. *Energy Policy*, 34(14), 1797-1806.
- Itaoka, K., Saito, A., & Akai, M. (2011). A study on roles of public survey and focus groups to assess public opinions for CCS implementation. [doi: 10.1016/j.egypro.2011.02.649]. *Energy Procedia*, 4(0), 6330-6337.
- King, W. R., & He, J. (2006). A meta-analysis of the technology acceptance model. [doi: 10.1016/j.im.2006.05.003]. *Information & amp; Management, 43*(6), 740-755.
- Krueger, R. A., & Casey, M. A. (2009). Focus Groups: A Practical Guide for Applied Research: Sage.
- Legris, P., Ingham, J., & Collerette, P. (2003). Why do people use information technology? A critical review of the technology acceptance model. [doi: 10.1016/S0378-7206(01)00143-4]. *Information & amp; Management, 40*(3), 191-204.
- Porter, C. E., & Donthu, N. (2006). Using the technology acceptance model to explain how attitudes determine Internet usage: The role of perceived access barriers and demographics. [doi: 10.1016/j.jbusres.2006.06.003]. Journal of Business Research, 59(9), 999-1007.
- Straub, D., Keil, M., & Brenner, W. (1997). Testing the technology acceptance model across cultures: A three country study. [doi: 10.1016/S0378-7206(97)00026-8]. *Information & amp; Management*, 33(1), 1-11.
- West, J., Bailey, I., & Winter, M. (2010). Renewable energy policy and public perceptions of renewable energy: A cultural theory approach. [doi: 10.1016/j.enpol.2010.05.024]. *Energy Policy*, 38(10), 5739-5748.
- Woersdorfer, J. S., & Kaus, W. (2011). Will nonowners follow pioneer consumers in the adoption of solar thermal systems? Empirical evidence for northwestern Germany. [doi: 10.1016/j.ecolecon.2011.04.005]. *Ecological Economics*, 70(12), 2282-2291.

Daily diary study of marital support in couples dealing with cancer

Aleksandra Gronostaj,

Institute of Psychology, Jagiellonian University, aleksandra.gronostaj@uj.edu.pl

Abstract. The aim of this study is to investigate gender and role (i.e. patient versus partner) differences in levels of stress and the outcomes of provided and received marital support in couples dealing with cancer. Colon cancer patients and their spouses are asked to concurrently complete a diary during a week before planned surgery and in a week after coming back home from the hospital, answering daily questions on mood, physical wellbeing, given and received support. Accounts from both spouses will allow to determine what actions are considered supportive, what are the differences in how patients and their partners report given and received help, whether there is a relationship between support and wellbeing and if there are gender differences in those aspects.

Keywords: social support, cancer, gender differences, diary study

1 Introduction

Life partner is traditionally considered a key source of support. In some research, marital status alone is a sufficient indicator of social resources in times of stress (Sęk, Cieślak, 2004). Support from a partner is linked to better adjustment to diseases, better mood, faster recovery and improved immunological markers in cancer patients (Hays, Sherbourne, Mazel, 1995; Manne, Pape, Taylor, Dougherty, 1999; Bodenmann, 1997; Kaniasty, 2005). At times though, it generates costs, and not only for the support provider, but surprisingly for the support receiver as well (Newsom, 1999; Fisher, Nadler, Whitcher-Alagna, 1982; Fry, 2001). Being helped can invoke feelings of indebtedness, dependency, inefficiency, inferiority. Conye, Ellard and Smith (1990) described for example 'helping dilemmas' of wives tending to their ill husbands, when attempts to help threatened the sense of self-rule already weakened by the disease. Men and women, when analyzing gender roles and socialization processes, seem to differ in terms of giving and receiving support (Miluska, 1996). It

is acceptable for a woman to both be the receiver and provider of help, whereas a man could be a support provider but not the one that is helped. To complicate matters even more some researchers reported counter intuitive findings that suggested that in happier marriages self-reported acts of spousal support were less frequent (Coyne, Bolger, 1990). A key to this paradox is perhaps in the hypothesis of invisible social support. Liberman (1986) pointed out that in good relationships helping may happen so smoothly and subtly, that it could go unnoticed. Unfortunately there is little research on support exchanges between spouses, especially in a situation of as severe a stress as cancer and surgery. This kind of crisis touches marital pair as a whole additionally complicating the support dynamics (Hagedoorn, Sanderman, Bolks, Tuinstra, Coyne, 2008). The aim of this study is to better describe it. Asking both partners about received and provided support will allow to determine: if there is a specific kind of supportive acts that go unnoticed more often (and are thus safer for self-esteem), if there are discrepancies between men and women in the amount and type of reportedly given and received assistance, how it interacts with their role as a patient or partner and what is the impact of help on their mood. Apart from scientific value, those questions are important for practical reasons, as it is still not clear how to give good support when dealing with cancer and surgery (Shiozaki, Hirai, Koyama, Inui, Yoshida, Tokoro, 2011). The information from this research about help dynamics and its outcomes in marital couples facing oncological disease could be utilized in designing psycho-educational and intervention programs for both spouses.

2 Literature review

2.1 Social support – a term of many meanings

Researchers talk of social support in so many contexts, that there were repeated calls for redefining the concept, so it could become a scientific one (Sęk, 1986). In the process of organizing the field, basics distinctions were made, and for example structural social support was proposed as a name for a net of connections – such as having a spouse, being a part of a church group, being able to name multiple friends. The most interesting division however was the one between perceived and received social support. The first one is measured by items like "There are people who offer me help when I need it." and tackles a belief that in times of stress help is going to come. The second is assessed by items like "[My spouse] took care of things I could not manage on my own during the last week." (Berlin Social Support Scales, Schultz, Schwartzer, 2003), and describes actual actions that took place in recent past. It turns out that perceived and received social support do not have much in common, most importantly have different impacts on mood. While the first one is invariably positive, the second generates costs.

2.2 Costs of receiving help

Supportive act as a social interaction inherently contains elements that could be burdening. Even if the help is needed and appropriate for the situation, there are at least four reasons why a person that receives it might feel bad (Schrout, Herman, Bolger, 2006; Gable, Algoe, 2010). First the asymmetrical relation between the helper and a person being helped may provoke negative social comparisons. Second the act of support may restrict the autonomy of the one already under stress. Third it may focus attention on the problem. Fourth, as the rule of reciprocity states, it provokes feelings of indebtedness, obligation to give back.

2.3 Invisible support

The term 'invisible support' was first used by Bolger, Zuckerman and Kessler (2000). Researchers asked married couples to report whether or not they have given or received support on a particular day. It was discovered that when one spouse reported giving support but the other reported getting none the mood of the latter was improved the most. How can one hide a supportive act so that it was helpful but not generating costs connected with help? There seem to be two ways to do so. One might deal with potential stressor before it becomes a problem for the other person, or the act of support might be disguised as something different in nature (e.g. doing ones duties). Unfortunately, because specific reports on what did the spouses do for each other are lacking, there is no way of telling what exactly an act of invisible support could look like.

2.4 Support and gender

Gender roles and stereotypes influence giving and receiving support in men and women. As active, independent and competent, men should be able to deal with problems on their own. A caring, passive and delicate woman can be a recipient of assistance on one hand and on the other is delegated to look after those in need. Research shows that the supportive networks of women are wide, while those of men are often restricted to the spouse (Oygard, 2001), that women join cancer support groups more often while men do it unwillingly (Volkers, 1999), that teenage boys shy from peer support to preserve their 'macho' image (Covie, Naylor, Chauhan, Smith, 2002). Accepting support seems to be especially difficult for men and threatening to manliness. Invisibility of help might be of greater importance in their cases.

2.5 Marriage, cancer and stress

Levels of stress in cancer patients are comparable irrespectively of cancer site ("Clinical practice guidelines for the psychosocial care of adults with cancer," 2003). In marital couples facing cancer, although the roles of patient and partner are clear and easily ascribed, the need for support and stress levels are not. It was suggested for example that being a woman and a partner to a cancer patient was as stressful as

actually suffering from the disease itself (Hagedoorn, Buunk, Kuijer, Wobbes, Sanderman, 2000). Thus the dynamics of support in marital pairs facing cancer may be especially complicated and provision and receipt of support may stray from obvious patterns from partner to patient, as cancer presents a crisis for marriage couple as a whole.

3 Proposed methodology

3.1 Research questions

In the context of reviewed literature this study will concentrate on three basic research questions:

- 1. Do men and women report different types and quantities of given and received support and does it depend on their role, i.e. partner or patient?
- 2. Are there specific kinds of supportive acts that go unnoticed more often?
- 3. Does provided and received support affect the mood of patients and partners and are there any gender differences?

3.2 Participants

60 married couples of colorectal cancer patients that await surgery and 60 married couples of colorectal cancer patients just after medical procedure will be recruited for the study. In half of the couples the wife and in half the husband will be the person suffering from cancer. The prognosis for colorectal cancer patients is relatively good, and male and female patients are roughly in the same age. Patients and their spouses are being recruited in five medical facilities in Cracow. The participation in the study is voluntary. Excluded are patients in palliative care, persons that suffer from psychiatric disorders or have difficulties in comprehending and correctly completing the questionnaires.

3.3 Procedure

Patients and their spouses are asked to complete three background questionnaires and a diary during the 7 days preceding the surgery or during 7 days following their release back home from hospital. Completed materials are returned to the hospital or sent directly to the researcher in included pre-stamped envelopes.

Small pilot study on 10 healthy couples and 5 couples facing cancer was conducted to test the procedure. Until this moment questionnaires from 19 couples before and 18 couples after surgery were collected. The current percentage of people that agree to take part in the project is 62%.

Study protocol was approved by Jagiellonian University Bioethics Committee. Approached patients and their spouses are informed that the participation is voluntary, anonymous, not connected to any medical procedure, and if agreed to, can be terminated at any moment.

3.4 Key variables and instruments

Given and received support. The diaries include two questions that participants answer daily over a period of one week: 'Did you support your partner today for a worry, problem or difficulty? If so, briefly describe how.' and 'Did your partner support you today for a worry, problem or difficulty? If so, briefly describe how.' Participants are instructed to answer these questions just how they understand them and are told that their subjective point of view is what matters. Reports will be analyzed by competent judges and supportive acts divided into categories in accordance with Curtona and Russell (1990) metaanalysis.

Mood. Adjectives from Profile of Mood States, three from Anger, three from Depression and three from Anxiety subscales were joined with Thayer's High Activation scale and positive adjectives to create balanced 18-item list. The list is included in the diary, and participants are asked to assess each evening the extent to which they were experiencing emotions described by the adjectives during the day. Ratings are on a 5-point scale ranging from 0 to 4.

Gender. Polish version of Bem Sex Role Inventory (Kuczyńska, 1992) – a list of 35 personality characteristics designed to assess masculinity, femininity and androgyny is one of background questionnaires. Participants are asked to rate themselves on a scale of 1 to 5 how much of a given personality characteristic they possess.

4 Conclusions

The data on specific support interactions and stress levels in couples dealing with cancer can contribute to better understanding of the proposed invisible support phenomenon, and more importantly, can provide insights into what types of supportive acts improve the well-being and which burden patients and partners the most. With that knowledge we'll be able to determine how to adequately support men and women, with respect to sometimes negative consequences of help.

References

Bodenmann, G. (1997) Dyadic coping—a systemic-transactional view of stress and coping among couples: Theory and empirical findings. *European Review of Applied Psychology*. 47, 2, 137-140

Bolger, N., Zuckerman, A., and Kessler, R.C. (2000) Invisible support and adjustment to stress. *Journal of Personality and Social Psychology*. 79, 6, 953-961

Clinical practice guidelines for the psychosocial care of adults with cancer (2003) National Breast Cancer Centre (Australia)

Coyne, J. C., Bolger, N. (1990) Doing without social support as an explanatory concept. *Journal of Social and Clinical Psychology*. 9, 1, 148-158

Conye, J. C., Ellard J. H., Smith D. A. F. (1990) Social support, interdependence and the dilemmas of helping, in Sarason, B. R., Sarason, I. G., Pierce, G. R. (eds), *Social support: An interactional view*, Willey, New York, 129-149

Covie, H., Naylor, P., Chauhan, L.T.P., Smith, P.K. (2002) Knowledge, use of and attitudes towards peer support: a 2 year follow-up to the Prince's Trust Survey. *Journal of Adolescence*. 25, 5, 453-467

Curtona, C. E., Russell, D. W. (1990) Type of social support and specific stress. Toward a theory of optimal matching, in Sarason, B. R., Sarason, I. G., Pierce, G. R. (eds), *Social support: An interactional view*, Willey, New York, 319-366

Fisher, J. D., Nadler, A., Whitcher-Alagna, S. (1982) Recipient reaction to aid. *Psychological Bulletin.* 91, 1, 27-54

Fry, P. S. (2001) Predictors of health-related quality of life perspectives, self-esteem, and life satisfaction of older adults following spousal loss: An 18-month follow-up study of widows and widowers. *The Gerontologist.* 41, 6, 787-798

Gable, S. L., Algoe, S. B. (2010) Being there when things go right: Support processes for positive events. in K. T. Sullivan, J. Davila (eds), *Support Processes in Intimate Relationships*, Oxford University Press, New York, 200-216

Hagedoorn, M., Buunk, B. P., Kuijer, R. G., Wobbes, Th., and Sanderman, R. (2000) Couples dealing with cancer: Role and gender differences regarding depressive symptoms and quality of life. *Psycho-oncology*. 9, 3, 232-242

Hagedoorn, M., Sanderman, R., Bolks, H. N., Tuinstra, J., Coyne, J. C. (2008) Distress in couples coping with cancer: A meta-analysis and critical review of role and gender effects. *Psychological Bulletin.* 134, 1, 1-30

Hays, R.D., Sherbourne, C.D., Mazel, R.M. (1995) http://www.rand.org/health/surveys_tools/mos/

Kaniasty, K. (2005) Social support and traumatic stress. *PTSD Research Quarterly*. 16, 2, 1-3

Kuczyńska, A. (1992) Inwentarz do oceny płci psychologicznej. Podręcznik, Pracownia Testów Psychologicznych PTP, Warsaw

Lieberman, M. A. (1986) Social supports – The consequences of psychologizing: A commentary. *Journal of Consulting and Clinical Psychology*. 54, 4, 461-465

Manne, S. L., Pape, St. J., Taylor, K. L., Dougherty, J. (1999) Spouse support, coping, and mood among individuals with cancer. *Annals of Behavioral Medicine*. 21, 2, 111–121

Miluska, J. (1996) *Tożsamość kobiet i mężczyzn w cyklu życia*, Wydawnictwo UAM, Poznań

Newsom, J. T. (1999) Another side of caregiving: Negative reactions to being helped. *Current Directions in Psychological Science*. 8, 6, 183-187

Oygard, L. (2001) Therapeutic factors in divorce support groups. *Journal of Divorce and Remarriage*. 36, 1-2, 141-158

Schrout, P. E., Herman, C. M., Bolger, N. (2006) The costs and benefits of practical and emotional support on adjustment: A daily diary study of couples experiencing acute stress. *Personal Relationships*. 13, 1, 115-134

Schulz, U., Schwarzer R. (2003) Social support in coping with illness: the Berlin Social Support Scales (BSSS). *Diagnostica*. 49, 2, 73-82

Shiozaki, M., Hirai, K, Koyama, A, Inui, H, Yoshida, R, Tokoro, A. (2011) Negative support of significant others affects psychological adjustment in breast cancer patients. *Psychology and Health.* 26, 11, 1540-1551

Sęk, H. (1986) Wsparcie społeczne – co zrobić aby stało się pojęciem naukowym? *Przegląd psychologiczny*. 29, 3, 791-800

Sęk, H., Cieślak, R. (2004) Wsparcie społeczne – sposoby definiowania, rodzaje i źródła wsparcia, wybrane koncepcje teoretyczne, In Sęk, H., Cieślak, R. (eds), *Wsparcie społeczne, stres i zdrowie*. Wydawnictwo Naukowe PWN, Warsaw, 11-28

Volkers, N. (1999) In coping with cancer, gender matters. *Journal of the National Cancer Institute*. 91, 20, 1712-1714

Better life indicators with Principal Component Analysis

Vasileios Ismyrlis

Phd Student, University of Macedonia, Department of Marketing and Operations Management-Edessa-Greece vasismir@gmail.com

Abstract. Studies to evaluate life satisfaction are now flourishing as there is available data from many sources. Many statistical analysis have been done and in this paper a method from the data analysis field is used. This method which reduces the number of variables, is Principal Component Analysis (PCA). Data from the Organization for Economic Co-operation and development (OECD) better life initiative is used with an effort to compare characteristics of the indicators (variables) and the countries. The present study examines the relationships between indicators of better life, and indicated four groups of them.

Keywords: Better life, Principal component analysis, Indicators, Relationships

1 Introduction – Literature review

Even if we can not define precisely well-being or in other words quality of life, because it has many aspects, as it is a concept with many contradictory opinions, we can say that it is every factor that influence people's everyday life. It is a multidimensional concept, including many aspects of everyday life, expressed as social indicators on observable outcomes in a variety of fields, such as leisure, family, literacy, health, employment etc.

Governments and other policy makers are trying to find ways to measure, express and compare the quality of people's living conditions in many different areas of everyday life. The reason to do this is to maintain people satisfaction in high level and plan their policies in the appropriate way to achieve it. There is a growing interest in the last years also expressed from economists who try to decode people's well-being preferences so as to explore individual perceptions. Consequently this part of social life is no longer solely the domain of social researchers. Other parties like NGOs and official statistical offices are getting more and more involved. The main issue is to assure reliable data and appropriate tools to measure well-being.

Well-being indicators are used to broadly illustrate people's general satisfaction with life, or give a more complete picture of quality of life in relation to many characteristics like their profession status, family life, health conditions, and standards of living.

Well-being measures can be both "subjective" and "objective". The subjective measures are based on individual's evaluation of their quality of life, usually extracted with the help of questionnaires, which make it possible to capture direct measures of high complexity concepts such as life-satisfaction (also referred as happiness).

Objective measures, on the other hand, attempt to capture these complex lifesatisfaction variables by looking at indicatory variables, such as leisure time, health status, disposable income and many others.

Strictly economic measures have failed to capture many of the social factors that influence people's lives, such as security, leisure, environment, etc. In many researches evidence exist that income is not the most significant factor of wellbeing (Sarah Fleche et al 2011). It is now fully recognized that to identify a better way to measure the progress of societies we have to move beyond GDP and examine more areas that impact everyday people's lives. The financial and economic crisis of the past few years is another factor that leads us to this conclusion.

There has been many initiatives from countries and organizations to measure wellbeing, like European Commission's "GDP and beyond", "Commission on the measurement of Economic Performance and Social Progress" and many others.

Such kind of study which tries to examine aspects of everyday life, is the better life initiative from OECD with the adoption of the better life index, which was launched in May 2011. OECD has expressed through the years an extended interest on quality of life and social indicators. The first step was in 1986 when OECD presented its social indicators. (OECD 1986).

The data used in the researches mostly come from official sources such as the OECD or National Accounts, United Nations Statistics, National Statistics Offices. A couple of indicators are based on data from the Gallup World Poll a division of the Gallup Organization that regularly conducts public opinion polls in more than 140 countries around the world.

The 11 topics of well-being used in the Index have been chosen in accordance with theory, practice and consultation on the issue of how to best measure well-being from a comparative perspective (<u>http://oecdbetterlifeindex.org/about/better-life-initiative/</u>). They can be separated in two categories (OECD 2011) as presented in Table 1.

Key topics	Indicators
Material living conditions	Housing, Income, Jobs
Quality of life	Community, Education, Environment, Governance, Health, Life satisfaction, Safety, Work-life balance

 Table 1 : Indicators' groups

The main aim of this paper is to analyze the data from the last better life report and to implement a method so as to find differences and compare the performance of the countries and to discover the variables-indicators' correlations. There was an effort to use a method from the field of the multidimensional statistical analysis, due to the existence of many dimensions.

2 Data and Methodology

OECD has 34 countries as members and most of them are developed economies. The data used, is from the 2011 OECD Better life initiative survey and includes information available until the year 2010.

From the original data of the 34 countries members (Table 2), and the 11 topics (including 25 indicators*) that exist in the index (Table 3), we have selected 28 countries, the OECD average and furthermore 7 topics (including 10 indicators with positive effect), with complete data for the countries included.

*the 2011 study included 20 indicators

ISO		Member	Developed	Included	GDP per
Code	Country	of EU*	**	***	capita****
AUS	Australia	Ν	Y	Y	40838
AUT	Austria	Y	Y	Y	41805
BEL	Belgium	Y	Y	Y	37677
CAN	Canada	Ν	Y	Y	40457
CHE	Switzerland	Y	Y	Y	43508
CHL	Chile	Ν	Ν	Y	16171
CZE	Czech	Y	Y	Y	25933
DEU	Deutschland	Y	Y	Y	37935
DNK	Denmark	Y	Y	Y	37741
ESP	Spain	Y	Y	Y	30622
EST	Estonia	Y	Y	Y	19375
FIN	Finland	Y	Y	Y	36723
FRA	France	Y	Y	Y	35048
GBR	United Kingdom	Y	Y	Y	35974
GRC	Greece	Y	Y	Y	27624
HUN	Hungary	Y	Ν	Y	19647
ICE	Iceland	Ν	Y	N	38079
IRL	Ireland	Y	Y	Y	39507
ISR	Israel	Ν	Y	N	31004
ITA	Italy	Y	Y	Y	30165
JPN	Japan	Y	Y	Y	34362
KOR	Korea	Y	Y	Y	31753
LUX	Luxemburg	Y	Y	N	84829
MEX	Mexico	Ν	Ν	N	15121
NLD	Netherlands	Y	Y	Y	42330
NZL	New Zealand	Y	Y	Y	27966
NOR	Norway	Y	Y	Y	53376
POL	Poland	Y	Y	Y	20136
PRT	Portugal	Y	Y	Y	23204
SVK	Slovak	Y	Y	Y	23384
SVN	Slovenia	Y	Y	Y	29179
SWE	Sweden	Y	Y	Y	40613
TUR	Turkey	Ν	Ν	N	14615
USA	United States	Ν	Y	N	48147
OECD	OECD	-	-	Y	-

 Table 2 : Countries-Members of OECD (the countries included in the analysis are in **bold**)

*Y=member of EU, N=not a member

**Y=Developed, N=Developing

***Y= included in the research, N=not included

**** in thousand dollars, Source: IMF 2011

Topics	Indicators	Pos/Neg*	Subj/Obj**	Included***
1.Housing	Rooms per person	+	0	Y
Housing	Dwelling without basic facilities	-	О	Ν
Housing	Household disposable income	+	О	Ν
Housing	Household financial wealth	+	О	Ν
2.Jobs	Employment rate	+	0	Y
Jobs	Long-term unemployment rate	-	О	Ν
3.Community	Quality of support network	+	0	Y
4.Education	Educational attainment	+	0	Y
Education	Students reading skills	+	0	Y
5.Environment	Air pollution	-	0	Ν
6.Governance	Voter turnout	+	0	Y
Governance	Consultation on rule-making	+	0	Y
7.Health	Life expectancy	+	0	Y
Health	Self-reported health	+	0	Y
8.Life Satisfaction	Life Satisfaction	+	S	Y
9.Safety	Homicide rate	-	0	Ν
Safety	Assault rate	-	0	Ν
10.Work-life balance	Employees working very long hours	-	0	Ν
Work-life balance	Employment rate of women with children	-	0	Ν
Work-life balance	Time devoted to leis- ure and personal care	_	О	Ν

Table 3 : The 10 topics-20 indicators of the 2011 data (the indicators included in the analysis are in **bold**)

*+=positive affect, -=negative **S=subjective indicator, O=objective ***Y= included in the research, N=not included

For the analysis of the data we used the Principal Component Analysis as the most suitable for discovering relationships, in cases of multidimensional data, between the variables (here indicators) and to define similar characteristics of the objects (here countries). PCA analysis is an exploratory technique of the Data Analysis field, which does not assume any distribution for the data and is putting forward possible relations that exist in the data, graphically as well. This can be achieved by the visualization of data in two or three dimension systems, maintaining their original structure. The method creates new principal components (also called factors) which are linearly uncorrelated. It can be used to determine how many real dimensions there are in the data, by replacing groups of the original variables with new, the components.

3 Analysis-Results

Data analysis and coding were processed with IBM SPSS (IBM SPSS, 2010) and M.A.D. Software (Carapistolis D., 2004).

Some elementary statistics of the original data (Table 14-Appendix) are now presented in Table 4.

	Range	Min.	Max.	Mean	Std. Dev.	Skew.	Kurt.
Rooms	1,5	1	2,5	1,67	0,43	0,12	-0,81
Employm	23,19	55,4	78,59	66,24	6,41	0,06	-1,19
Support	17,5	79,8	97,3	91,45	4,54	-0,87	0,07
Education	62,65	28,25	90,9	74,98	13,70	-1,63	3,69
Reading	90	449	539	497,66	19,25	0,06	0,82
Cons_rule	9,5	2	11,5	7,54	2,53	-0,33	-0,57
Voter	47	48	95	72,10	11,87	-0,02	-0,73
Life_expect	8,9	73,8	82,7	79,39	2,31	-1,26	1,03
Health	58,6	31,1	89,7	67,46	15,27	-0,82	0,27
Life_satisf	3,1	4,7	7,8	6,64	0,86	-0,65	-0,29

Table 4 : Descriptive statistics

The results of the Principal Component Analysis are following:

	Roo ms per per- son	Em- ploy- ment rate	Sup- port net- work	Edu- ca- tional	Read ing	Cons on rule mak- ing	Voter turno ut	Life ex- pect- ancy	Self- sup- por- ted healt h	Life satis- fac- tion
Room s per per- son	1	0,566	0,683	_ 0,105	0,455	0,212	0,464	0,648	0,679	0,747
Em- ploy- ment rate	0,566	1	0,527	0,271	0,366	0,331	0,129	0,547	0,457	0,699
Sup- port net- work	0,683	0,527	1	0,216	0,138	0,357	0,267	0,349	0,682	0,74
Edu- ca- tional	-0,11	0,271	0,216	1	0,223	0,197	_ 0,305	-0,26	- 0,066	0,217
Read- ing	0,455	0,366	0,138	0,223	1	0,474	- 0,086	0,28	0,144	0,24
Cons on rule	0,212	0,331	0,357	0,197	0,474	1	- 0,326	0,182	0,252	0,263
Voter turno ut	0,464	0,129	0,267	_ 0,305	- 0,086	_ 0,326	1	0,369	0,369	0,403
Life ex- pect-	0,648	0,547	0,349	-0,26	0,28	0,182	0,369	1	0,425	0,603
Self- sup- por- ted healt h	0,679	0,457	0,682	0,066	0,144	0,252	0,369	0,425	1	0,687
Life satis- fac-	0,747	0,699	0,74	0,217	0,24	0,263	0,403	0,603	0,687	1

 Table 5 : Observed correlation matrix

We can observe (Table 5), that the correlation coefficients between the variables have interesting values and are adequate for further research.

For example the indicator "Rooms per person" has seven positive correlations (with values over 0.3), which indicates that has the same attributes with the most of the indicators. "Voter turnout" on the other hand has two negative correlations and two positives and seems not to be strongly correlated with any of the variables.

	Table 6	KMO a	nd Bartlett's	Test
--	---------	-------	---------------	------

Kaiser-Meyer-Olkin Measure	,693	
Bartlett's Test of Sphericity	Approx. Chi-Square	157,401
	df	45
	Sig.	,000

The value of KMO test which is 0.693 and is over 0.50 and the value of Sig which is zero (Table 6), indicate that there are significant correlations among some variables.

Table 7 : Communalities

Variable	Initial	Extraction
Rooms_per_person	1,000	,846
Employment	1,000	,647
Support_network	1,000	,788
Educational	1,000	,814
Reading	1,000	,719
Cons_on_rule_making	1,000	,671
Voter_turnout	1,000	,721
Life_expectancy	1,000	,737
Self_supported_health	1,000	,668
Life_satisfaction	1,000	,875

The value "Extraction" from Table 7, indicates the proportion of each variable's variance, explained by the retained components. Value 1 indicates that components explain 100% of the variance. In our case there are satisfactory values (over 0.60) of "Extraction" in all the variables, so we can conclude that all the variables are well represented in the common component space.

The unrotated component loadings, which indicate how closely the variables are related to each component, are presented below (Table 8). These numbers show that on the first component we have six high positive loadings*. The variables "Rooms per person", "Employment", "Support network", "Life expectancy", "Self Supported Health", and "Life satisfaction" have loaded highly on component 1.

		Component	
	1	2	3
Rooms per person	,893	-,165	-,148
Employment	,767	,240	,047
Support_network	,802	,067	,375
Educational	,086	,700	,562
Reading	,433	,504	-,527
Cons_on_rule	,396	,661	-,278
Voter_turnout	,412	-,736	,102
Life_expectancy	,711	-,247	-,412
Self_supported_health	,784	-,167	,158
Life_satisfaction	,905	-,022	,235

Table 8 : Component matrix

The variables "Educational", "Reading", "Cons_on_rule", and "Voter_turnout", have loaded highly on component 2 and the variables "Educational" and "Reading" on component 3.

*highly loading is defined when there are values over 0.50 or -0.50

Table 9 : Rotated Component Matrix

		Component	
	1	2	3
Rooms per person	,806	,274	,349
Employment	,697	,393	-,089
Support_network	,864	,076	-,192
Educational	,171	,162	-,871
Reading	,157	,829	,078
Cons_on_rule	,184	,770	-,211
Voter turnout	,521	-,443	,504
Life expectancy	,560	,324	,564
Self_supported_health	,807	,043	,122
Life_satisfaction	,925	,137	-,016

Table 9, shows the component loadings that result after the Varimax rotation. The component loadings indicate how closely the variables are related to each component.

The variable belongs to the component with the greater loading (in **bold**). So the variables "Rooms per person", "Employment", "Support network", "Voter turnout", "Self supported health", and "Life_satisfaction" belong to component 1, the "Reading" and "Cons_on_rule" to component 2 and "Education" and "Life expectancy" to component 3.

Therefore it can be seen that variables "Voter_turnout" and "Life_expectancy" have almost the same values to both the 1 and 3 components

From the above table, we can measure the internal consistency of the first factor (which includes six of the variables), with the help of the Cronbach's Alpha coefficient which has a satisfactory value of 0.617. For the second factor we have only two loadings and for the third only one positive loading and the Cronbach 's Alpha is not applicable.

Table 10 : (Component	Transformation	Matrix
--------------	-----------	----------------	--------

Component	1	2	3
1	,932	,332	,146
2	-,139	,699	-,701
3	,334	-,633	-,698

Table 10, is the matrix by which you multiply the unrotated component matrix (Table 8) to get the rotated component matrix (Table 9). It gives information about the extent to which the components have been rotated.

6.078

3,512

2,919

2,550

1,377

.843

88,799

92,311

95,230

97,780

99,157

100.000

		-		
				Extraction Sums of
Component		Initial Eigenvalu	ies	Squared Loadings
	Total	% of Variance	Cumulative %	Total
1	4,490	44,902	44,902	4,490
2	1,901	19,009	63,911	1,901
3	1,096	10,959	74,870	1,096
4	,785	7,851	82,721	

Table 11a : Total Variance Explained

,608

,351

,292

.255

.138 .084

5

6 7

8

9

10

Table 11b : Total Variance Explained

Componen	Extraction Sur	ns of Squared						
t	Load	ings	Rotatio	Rotation Sums of Squared Loadings				
	% of Variance	Cumulative %	Total	% of Variance	Cumulative %			
1	44,902	44,902	4,060	40,602	40,602			
2	19,009	63,911	1,863	18,633	59,235			
3	10,959	74,870	1,564	15,635	74,870			
4								

The two tables above (11a,11b) show the variances explained by the components.

The initial number of components (column Component-Table 11a) is the same as the number of variables used in PCA. The Table 11a also shows the importance of the 10 principal components. The first three components have eigenvalues over 1.00 accounting for 74.87% of the total variance of the data.

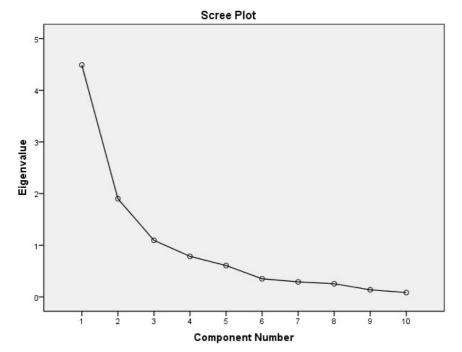
From the above Tables 11a,11b we can conclude that the three components explain with both the methods (Extraction Sums of Squared Loadings, Rotation Sums of Squared Loadings) the same percentage of variance.

This leads to the conclusion that a three component solution will probably be adequate. In summary, it can be concluded that PCA has identified three components from the list of the ten variables.

- The first factorial axis (first principal component) interprets with a percentage of 44.90 the researched issue.
- The second factorial axis (second principal component) has a 19.01 interpretation percentage.
- The third factorial axis (third principal component) has a 10.96 interpretation percentage.

The first factorial space interprets data with a percentage of 63.91. The total information that the fist three axes provide, which are of interest in the present study, is 74.87%.

Figure 1 : Scree plot of eigenvalues



The scree plot above (Figure 1), graphs the eigenvalue against the component number.

From the scree plot of the eigenvalues, the Kaiser criterion (retain factors with eigenvalues greater than 1) and the analysis precedent we can conclude that the three components solution is an appropriate one.

The interpretation of axes that follow, is done firstly from individuals, secondly from variables.

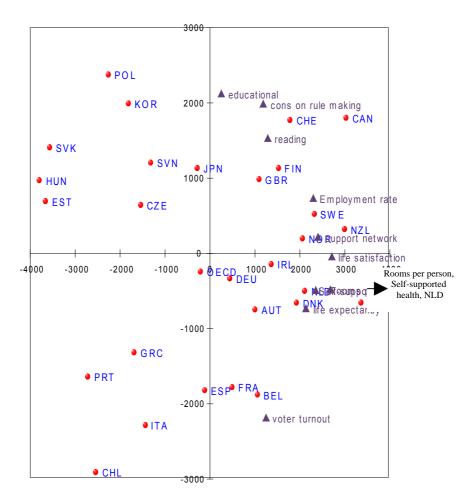
 Table 12 : Interpretation indicators: coordinates (#F), correlations (COR), contributions (CTR) of the objects-countries.

	#F1	COR	CTR	#F2	COR	CTR	#F3	COR	CTR
AUS	626	796	3	-124	31	0	-121	30	0
AUT	188	199	0	-141	111	0	257	373	2
BEL	198	148	0	-351	459	2	-3	0	0
CAN	564	634	2	332	220	2	-59	7	0
CHL	-470	344	2	-542	457	5	272	116	2
CZE	-284	441	1	117	76	0	249	341	2
DNK	359	550	1	-124	65	0	207	182	1
EST	-677	705	4	127	25	0	152	36	1
FIN	285	388	1	208	207	1	-134	84	1
FRA	94	52	0	-332	642	2	63	24	0
DEU	84	63	0	-64	36	0	154	210	1
GRC	-311	392	1	-247	246	1	-107	46	0
HUN	-703	770	4	178	50	1	79	10	0
IRL	255	296	1	-29	4	0	35	6	0
ITA	-264	209	1	-426	543	3	-202	121	1
JPN	-51	7	0	208	116	1	-252	169	2
KOR	-335	183	1	368	223	2	-491	395	8
NLD	393	742	1	-95	42	0	56	16	0
NZL	558	842	2	57	9	0	-90	22	0
NOR	385	809	1	35	7	0	65	23	0
POL	-417	387	1	439	430	3	90	18	0
PRT	-503	329	2	-306	122	2	-523	356	9
SVK	-660	677	3	259	105	1	253	100	2
SVN	-243	315	0	222	266	1	57	18	0
ESP	-20	2	0	-340	475	2	-140	80	1
SWE	434	715	1	95	34	0	101	39	0
CHE	333	257	1	327	248	2	40	4	0
GBR	205	241	0	181	188	1	-45	11	0
OEC	-37	196	0	-47	322	0	19	54	0

 Table 13 : Interpretation indicators: coordinates (#G), correlations (COR), contributions (CTR) of the variables-indicators.

	#G1	COR	CTR	#G2	COR	CTR	#G3	COR	CTR
Rooms per person	892	797	178	-165	27	14	-147	22	20
Employment rate	766	587	131	240	58	31	45	2	2
Support network	801	643	143	68	5	2	376	142	130
Educational	84	7	2	703	496	260	556	310	283
Reading	429	184	41	506	256	135	-532	283	258
Cons on rule	395	156	35	658	434	228	-282	79	72
Voter turnout	416	173	39	-733	536	282	94	9	8
Life expectancy	711	506	113	-249	62	33	-410	168	153
Self-supported	784	616	137	-168	28	15	161	26	24
Life satisfaction	904	819	182	-19	0	0	234	55	50

Figure 2: First factorial space



The first factorial space (Figure 2) displays simultaneously the first two components. It was created before the rotation of components. It explains a 63.91% of the variance and with the help of tables 12,13 we can assume the following:

- The first axis opposes the indicators "Rooms per person" and "Life satisfaction" to "Educational".
- The second axis opposes the indicators "Educational" and "Cons on rule making" to "Voter turnout".
- The third axis opposes the indicator "Educational" to "Reading" and "Life expectancy".
- The first axis opposes the countries AUS,CAN,NZL which have the same characteristics according to indicators, to the countries HUN,EST,SVK.
- The second axis opposes the countries CHL, ITA, BEL to the countries POL, KOR, CHE.
- The third axis opposes the countries PRT, KOR to the countries CHL, AUT.

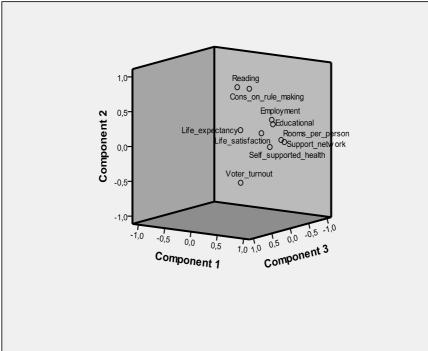


Figure 3 : Component plot

Component Plot in Rotated Space

SPSS has produced a plot (Figure 3) of the ten variables-indicators on axes (rotated component space) representing the three rotated components.

From this plot and the previous results, it can be concluded that variables-indicators, can be grouped due to their correlations and their representation in the rotated space, as follows:

"Voter turnout" has expressed different behaviour from all the other variables.

"Educational" has also a different attitude from all the other variables.

"Reading" and "Cons on rule making" are correlated with each other .

All the rest variables are close and highly related to each other and not with the above four indicators.

From the previous analysis, we can assume that there is differentiation detected for some indicators like "Voter turnout", "Educational", "Reading", "Cons on rule making", which seem not to be correlated with other indicators. Specifically, "Voter turnout" and "Educational", seem to be negative correlated with many of the indicators.

4 Conclusions

4.1 Implications

The general assumption that we could make from our research, is that from the 10 indicators of the index and the 28 countries-members of the OECD analysed in the current research, there were quite differences detected.

Countries seem to have different characteristics and for example the developing HUN, CHL or those with lower GDP, like GRC, PRT, ESP have lower score in the most indicators. Countries from specific regions, like North Europe (SWE, FIN DEN,) Central Europe (DEU, AUT, CHE), South Europe (GRC, PRT, ESP), East Europe (EST,HUN,SVK) seem to have similar characteristics.

From the indicators we have distinguished four of them that are apart and not correlated with all the others: "Voter turnout", "Educational", "Reading", "Cons on rule making".

4.2 Limitations-suggestions

A proposition for future research is to use data from another countries, and if possible more indicators and sources, like World value survey, Human development index and others.

A bigger size of sample and ratio of countries to indicators, could minimize the probability of errors and increase the generalization ability of the results.

References

Better life index, [Online] Available: <u>http://oecdbetterlifeindex.org/about/better-life-initiative/, [accessed 10-9-11]</u>

Beyond GDP, [Online] Available: <u>http://www.beyond-gdp.eu/indicatorList.html?</u> <u>indicator=Well-being</u>, [assessed 12-3-12]

Carapistolis D., (2004), M.A.D. Software.

Fleche Sarah S., Smith C., Sorsa P. (2011), Exploring determinants of subjective Wellbeing in OECD countries-Evidence form the World Value Survey, *OECD-Economics Department Working Papers*

IBM SPSS (2010) IBM SPSS Statistics 19 Core System User's Guide.

OECD (1986), Living Conditions in OECD Countries – A Compendium of Social Indicators, OECD, Paris.

OECD (2010), Society at a Glance 2009, OECD, Paris

OECD (2011), How's life? Measuring well-being 2011, OECD, Paris

Appendix

Table 14: Original data

	Room s per perso n	Empl oyme nt rate	Supp ort netwo rk	Educ ationa 1	Readi ng	Cons on rule maki ng	Voter turno ut	Life expec tancy	Self- suppo rted health	Life satisfa ction
AUS	2,40	72,30	95,40	69,72	514,90	10,50	95,17	81,50	84,90	7,50
AUT	1,70	71,73	94,60	81,04	470,28	7,13	81,71	80,50	69,60	7,30
BEL	2,30	62,01	92,60	69,58	505,95	4,50	91,08	79,80	76,70	6,90
CAN	2,50	71,68	95,30	87,07	524,24	10,50	59,52	80,70	88,10	7,70
CHL	1,30	59,32	85,20	67,97	449,37	2,00	87,67	77,80	56,20	6,60
CZE	1,30	65,00	88,90	90,90	478,19	6,75	64,47	77,30	68,20	6,20
DNK	1,90	73,44	96,80	74,56	494,92	7,00	86,59	78,80	74,30	7,80
EST	1,20	61,02	84,60	88,48	500,96	3,25	61,91	73,90	56,30	5,10
FIN	1,90	68,15	93,40	81,07	535,88	9,00	74,05	79,90	67,70	7,40
FRA	1,80	63,99	93,90	69,96	495,62	3,50	83,97	81,00	72,40	6,80
DEU	1,70	71,10	93,50	85,33	497,31	4,50	77,65	80,20	64,70	6,70
GRC	1,20	59,55	86,10	61,07	482,78	6,50	74,14	80,00	76,40	5,80
HUN	1,00	55,40	88,60	79,70	494,18	7,88	64,39	73,80	55,20	4,70
IRL	2,10	59,96	97,30	69,45	495,64	9,00	67,03	79,90	84,40	7,30
ITA	1,40	56,89	86,00	53,31	486,05	5,00	80,54	81,50	63,40	6,40
JPN	1,80	70,11	89,70	87,00	519,86	7,25	67,46	82,70	32,70	6,10
KOR	1,30	63,31	79,80	79,14	539,27	10,38	63,01	79,90	43,70	6,10
NLD	2,00	74,67	94,80	73,29	508,40	6,13	80,35	80,20	80,60	7,50
NZL	2,30	72,34	97,10	72,05	520,88	10,25	79,46	80,40	89,70	7,20
NOR	1,90	75,31	93,10	80,70	503,23	8,13	77,44	80,60	80,00	7,60
POL	1,00	59,26	92,20	87,15	500,48	10,75	53,88	75,60	57,70	5,80
PRT	1,50	65,55	83,30	28,25	489,33	6,50	64,26	79,30	48,60	4,90
SVK	1,10	58,76	89,60	89,93	477,44	6,63	54,67	74,80	31,10	6,10
SVN	1,10	66,20	90,70	82,04	483,08	10,25	63,10	78,80	58,80	6,10
ESP	1,90	58,55	94,10	51,23	481,04	7,25	75,32	81,20	69,80	6,20
SWE	1,80	72,73	96,20	85,04	497,45	10,88	81,99	81,20	79,10	7,50
CHE	1,70	78,59	93,20	86,81	500,50	8,38	48,28	82,20	80,95	7,50
GBR	1,80	69,51	94,90	69,63	494,18	11,50	61,36	79,70	76,00	7,00
OEC D	1,63	64,52	91,10	72,95	493,45	7,28	72,29	79,20	69,00	6,70

An Econometric and Social Analysis of the Impact of Tobacco Control Policies in the Hospitality and Tourism Industry in Cyprus

Christos Savva¹, Chrystalla Kapetaniou², Elpidoforos Soteriades³, and Lambros Lazouras⁴

¹ Department of Commerce Finance and Shipping, Cyprus University of Technology, Limassol, Cyprus

² Department of Commerce Finance and Shipping, Cyprus University of Technology, Limassol, Cyprus, Email: chrysta.kapetaniou@cut.ac.cy

³ Department of Environmental Health, Environmental and Occupational Medicine and

Epidemiology (EOME), Harvard School of Public Health, Boston, MA, USA

⁴ CITY College International Faculty of the University of Sheffield, Thessaloniki, Greece

Abstract. As of January 1st, 2010 Cyprus introduced a new law to ban smoking in public settings, and this may directly impact the financial revenues of the hospitality and tourism industry. On the other hand, the smoke-free law will significantly improve the quality of services offered to local customers and tourists, by protecting their right for smoke-free air. These two possibilities of the smoking ban raise questions on whether the new law will actually hurt or benefit the local hospitality and tourism business. The proposed project aims to provide scientific evidence to resolve this question by examining the empirical evidence on the actual and perceived financial impact of total smoking bans in the hospitality and tourism industry in Cyprus. The methodology employs two different approaches. The first approach concerns an econometric analysis of the costs and benefits of the smoking bans on businesses. The second approach concerns a cross-sectional survey study in a representative sample of managers and staff employed in hotels, restaurants, and bars/clubs in Cyprus. The findings of the proposed project are expected to generate important knowledge and help policy makers identify foreseeable financial risks and benefits of important public health decisions on the Cypriot economy. It will also help in promoting the realistic and positive economic effects of smoking

bans among professionals and stakeholders involved with the hospitality and tourism industry in the country.

Keywords: tobacco control policies, Cyprus, hospitality industry

1 Introduction

Being committed to reduce mortality and disease related to tobacco use, the World Health Organization (WHO) introduced the Framework Convention on Tobacco Control (FCTC), which has been signed and ratified by more than 130 countries, and mandates that effective action should be taken by signatory countries to reduce tobacco use in the population, as well as exposure to second hand smoke (SHS). Complying with FCTC regulations, the Cypriot government introduced a smoking ban in public places and workplace settings in January 1st, 2010. This has been an important step towards reducing exposure to passive smoking and de-normalizing tobacco in a country that tops the list of European countries with the highest smoking rates. Yet, the new smoking ban creates a dilemma for people working in the hospitality and tourism industry. Specifically, a counterargument against smoking bans in public places is that restricting smoking in hotels, restaurants and other typical leisure settings, will result in significant losses in clients, and, as a result, negatively impact economic growth.

Previous studies in other countries have consistently shown that smoking bans in key hospitality settings such as restaurants, bars and clubs, and hotels do not hurt hospitality and tourism business. However, no relevant studies have been conducted so far on the economic impact of smoke-free legislation on the hospitality/tourism industry in countries where smoking rates are still comparably high, and social norms against smoking are rather weak, such as Cyprus. The current proposed research project is aiming at evaluating the expected impact of the upcoming smoke-free policy at the very beginning in order to provide reliable scientific evidence on the potential positive or negative consequences of the above policy on the revenues of the hospitality industry.

2 Literature Review

Over time, a plethora of research has identified the health effects of smoking, and has now been widely accepted that not only primary smoking, but also second-hand smoke (SHS) is damaging to health. As the evidence of the health effects of smoking amassed, arguments over smoke-free legislation became increasingly focused on their economic impact. Both the elimination of second-hand smoking and the reduction in the amount of cigarettes consumed by workers and/ or customers create a healthier workforce, which in turn result in less compensation premiums and insurance payments, in improved productivity and attendance and in reduced health care costs. However, the effect of smoke-free regulation on the incidence of smoking is not straightforward. Some studies have noted that smoke-free policies do not reduce quitting directly, and in some cases, smokers oppose the policies and continue to smoke. Even more so, in countries like Greece where smoking is also high prevalent in the general population, smokers may completely disregard the health effects of SHS exposure, violate existing smoke-free policies, and smoke in front of non-smokers and children.

Accompanying the growth in smoke-free policies has been a parallel increase in allegations that smoke-free laws will hurt local economies and businesses. Though numerous studies have been conducted to assess the economic impact of smoke-free legislation, the quality of most studies varies greatly with many prior studies of smoke-free policies riddled with methodological flaws. Evidence from the best quality studies consistently demonstrates that a smoke-free policy does not have a negative impact on hospitality facilities over the long term. However, the smoke-free legislation cannot yield benefits with no costs. The successful realisation of benefits depends on widespread and uniform compliance. Enforcement of the tobacco control policies is necessary to ensure high level of compliance for its successful implementation.

3 Methodology

The methodology employs two different approaches. The first approach concerns an econometric analysis of the costs and benefits of the smoking bans on businesses. More specifically, it will focus on the revenue analysis of bars/restaurants before and after the imposition of highly restrictive restaurant smoking policies. The effect of restaurant smoking policies will be estimated using a fixed effects regression model, entering a panel of 24 months of data for the 5 districts of Cyprus. The econometric analysis will provide quantitative indications on the change in the trend in revenues following the implementation of highly restrictive restaurant smoking policies. Estimation processes will be performed using the appropriate econometric software.

The second approach concerns a cross-sectional survey study in a representative sample of managers and staff employed in hotels, restaurants, and bars/clubs in Cyprus. The survey study consists of three different parts. The first part involves a pilot study of the questionnaire to be used to make sure participants understand the questions, and, most importantly, that the questions elicit proper responses. Reliability analysis and validity assumptions will be examined at this stage following proper analysis of the pilot study's data. Depending of the pilot study findings, the main survey study will use either the same questionnaire or a modified version to collect data from a large and representative sample of managers/employers and employees working in key hospitality and tourism businesses. The final part of the survey study concerns the analysis of the obtained data. This will, first of all, describe the response rates, which are crucial for the validity and the potential for generalization of the project findings. Also, the analysis of the samoke-free law, and identify the predictors of willingness to support the ban, after controlling/adjusting for the effects of important covariates (e.g., smoking status, knowledge and perceived health

risks of passive smoking). For this purpose, both descriptive and multivariate analyses will be used.

4 Conclusions

Over the years, a plethora of studies has recognised the health problems caused by active smoking as well as passive smoking. This evidence has prompted a number of countries, including Cyprus, to design and implement effective and comprehensive tobacco control policies to reduce smoking prevalence and second-hand smoke. Cyprus smoke-free law signals an important step towards reducing tobacco use in the population, as well as exposure to second hand smoke (SHS) in a country that tops the list of European countries with the highest smoking rates.

The new smoking ban in public settings in Cyprus indicates the country's responsiveness to important societal issues and the ability to adapt to European-wide initiatives. Yet, following the worldwide economic recession, professionals in the hospitality and tourism industry are concerned about the economic side-effects of the ban. Despite the large and diverse body of literature on the economic effects of smoke-free policies, there's no evidence to guarantee that such a transition is accompanied by positive effects (or lack of negative effects) on other levels, such as the local economy in a country where smoking rates are significantly high, and social norms against tobacco use are comparatively weak.

The smooth transition from a smoking culture to a smoke-free policy requires a better understanding of the actual resultant financial risks and benefits, as well as a thorough understanding of interested parties and stakeholders' subjective estimates of those risks and benefits. The proposed project aims to provide timely and necessary feedback to the upcoming public health initiative, and, accordingly, promote informed decision-making.

References

Adda, J., & Cornaglia, F. (2010). The Effect of Taxes and Bans on Passive Smoking. *The American Economic Journal: Applied Economics, 2,* 1-32.

Allwright, S., Paul, G., Greiner, B., Mullally, B. J., Pursell, L., Kelly, A., Perry, I. J. (2005). Legislation for smoke-free workplaces and health of bar workers in Ireland: Before and after study. *British Medical Journal*, *331*, 1117.

Bauer, J. E., Hyland, A., Li, Q., Steger, C., & Cummings, K. M. (2005). A longitudinal assessment of the impact of smoke-free worksite policies on tobacco use. *American Journal of Public Health*, *95*(6), 1024-1029.

Buddelmeyer, H., & Wilkins, R. (2005). The effects of smoking ban regulations on individual smoking rates. *University of Melbourne: Discussion Paper No. 173.*

California Environmental Protection Agency, Office of Environmental Health Hazard Assessment. (1997). *Health effects of exposure to environmental tobacco smoke*. Sacramento, CA: Author.

Eadie, D., Heim, D., MacAskill, S., Ross, A., Hastings, G., & Davies, J. (2008). A Qualitative Analysis of Compliance with Smoke-Free Legislation in Community Bars in Scotland: Implications for Public Health, *Addiction*, *103* (6), 1019-1026.

Eisner, M. D., Smith, A. K., & Blanc, P. D. (1998). Bartenders' respiratory health after establishment of smoke-free bars and taverns. *Journal of the American Medical Association*, 9 (22), 1909-1914.

Glantz, S. A., & Charlesworth, A. (1999). Tourism and hotel revenues before and after passage of smoke-free restaurant ordinances. *Journal of the American Medical Association, 281,* 1911-1918.

Glantz, S. A., & Smith, L. R. (1997). The effects of ordinances requiring smoke-free restaurants and bars on revenues: A follow-up. *American Journal of Public Health*, *87*, 1687-1693.

Halpern, M. T., Shikiar, R., Rentz, A. M., & Khan, Z. M. (2001). Impact of smoking status on workplace absenteeism and productivity. *Tobacco Control*, 10 (3), 233-238.

Lazuras, L., Rodafinos, A., & Eiser, J. R. (2009). Predicting smokers' non-compliance with smoking restrictions in public places. *Tobacco Control, 18*, 127-131.

Parrott, S., Godfrey, C., Raw, M., West, R., & McNeill, A. (1998). Guidance for commissioners on the cost effectiveness of smoking cessation interventions. Thorax, 53 (suppl. 5, pt 2): S4–24.

Scollo, M., & Lal, A. (2008). Summary of studies assessing the economic impact of smoke-free policies in the hospitality industry – includes studies produced to January 2008. Melbourne: Vic Health Centre for Tobacco Control.

Scollo, M., Lal, A., Hyland, A., & Glantz, S. A. (2003). Review of the quality of studies on the economic effects of smoke-free policies on the hospitality industry. *Tobacco Control, 12,* 13-20.

TNS Opinion & Social (2010). *Special Eurobarometer 332- Tobacco*. Retrieved from http://ec.europa.eu/public_opinion/archives/ebs/ebs_332_sum_en.pdf.

Wakefield, M. A., Wilson, D., Owen, N., Esterman, A., & Roberts, L. (1992). Workplace smoking restrictions, occupational status, and reduced cigarette consumption. *Journal of Occupational Medicine*, *34* (7), 693-697.

Motivation of Students at the Palacký University in Olomouc for Their Specialization in Teacher Profession

Pavla Nádvorníková

Institute of Education and Social Studies, Faculty of Education, Palacký University in Olomouc, P.Nadvornikova@seznam.cz

Abstract. The topic "Motivation of Students at the Palacký University in Olomouc for Their Specialization in Teacher Profession" is based on psychology and pedagogy, its results and application belong to the field of pedagogy. The main aim of the thesis is to identify motivation factors, which influence motivation of students at the Faculty of Education at the Palacký University in Olomouc for their specialization in teacher profession, and to compare them with motivation of students at other faculties at the Palacký University in Olomouc. The paper deals with the methodology of the thesis and theoretical basis subject of the post.

Keywords: *motivation, motivation factors, study, teacher profession*

1 Introduction

Motivation is very important in human life. In the educational process, it is understood in two ways (Hrabal, Man, Pavelková, 1989): a) as means of increasing the efficiency of learning activities of students, and b) as one of the important educational goals of the school.

My thesis deals with studentsø motivation to preparation for teacher profession. In the theoretical part I will focus on basic theoretical terminology such as motivation, pupil (student), characteristics and study objectives of pedagogical programmes etc. The aim of the thesis is to identify studentsø motivation for their studies at faculties of education. The practical part will be based mainly on the quantitative methods with a questionnaire on the top. The research is focused only on the Czech Republic, in particular only on the Palacký University in Olomouc.

When choosing the field of study a student is influenced by many opinions and facts. I expect to find out what portion of students is studying teaching subjects only provisionally and regards it as a "transfer station" for further study, how many students are enrolled in studying teaching subjects due to parents' wishes, or how many students wants to really become a teacher after graduation.

The article is a description of the principal terms, which will be a part of a theoretic part of the thesis. The main aim of this paper is to characterize term motivation and to describe methodology of research.

2 Motivation

There is no discussion on the fact, that motivation plays important role in a personal life. Nakone ný (1996, p. 17) defines motivation as õintrapsychic process originating from the internal and external situation of the individualö. The topic of motivation is not clearly defined within the field of psychology. There are many approaches towards motivation. The first group focuses on the content of motivation, i.e. what motivates an individual, what are the basic motivations, and what are their mutual relations. The other group is interested in the process of motivation, i.e. how the motivations influence individualøs behaviour (Lok–a, Lok–ová, 1999, p. 10). Nakone ný (1973, p. 106) mentions two different attitudes in motivation concept according to the preferred basic component of the behaviour: a) focused on specific objective and b) focused on specific level of energy.

Zelinková (2007, p. 125) defines motivation as an aggregate of driving moments in activities, experience, behaviour and personalities. The driving moment is what leads us to certain behaviour or what discourage us from certain behaviour. Motivation is the reason why people act and behave in certain way.

Motivation in pedagogical educational process is understood in two ways (Hrabal, Man, Pavelková, 1989): a) as means of increasing efficiency of learning activities of students (questions of motivating the students in classes) and b) as one of the important pedagogical educational goals of the school (questions of development of motivational sphere of students). These authors mention that if we understand learning activity as one of the important forms of cognitive activity happening at school, we may think about at least three sources of motivation of this activity. The first aspect is related to cognition and acquisition of new knowledge ó cognition needs, the second aspect is related to social relations ó social needs, and the third one related to level of tasksø difficulty ó performance needs. The authors renewed their list (Hrabal, Pavelková, 2010, p 121) by adding needs of responsibility and of being in accordance with moral values ó when we arouse in students responsibility for their own results and responsibilities to fulfil schooløs demands, and by adding perspective orientation ó needs to outline their own future.

2.1 Theory of motivation

The process of motivation and creation of motivation are dealt by the theories of motivation and motivation process. They explain why the people behave in a certain manner. Let simply mention the principles of some theories of motivation. Each theory approaches the issue from a different point of view, and each of them has its own substantiation for it (Armstrong, 2007).

The fundamental prerequisite of the *theory of instrumentality* is that the individual will be motivated to work if the rewards and punishments are directly linked with his or her performance, so that the rewards depend on his or her actual performance. The instrumentality is an idea that discharge of one activity will lead to discharge of another one (Armstrong, 2007).

Another theory of motivation falling within the field of management of resources is the *Herzberg Two-Factor Theory*, whose author is Frederick Herzberg, who formulated it in 1959. He addressed the two basic factors that form the resource of satisfaction and motivation. The first of them are hygienic factors, which are those that cause dissatisfaction (working conditions, interpersonal relations). The second factor is motivators that help to raise the satisfaction and motivation. The fulfilment of hygienic factors is considered as a matter of course and its non-fulfilment causes anger. On the contrary, the fulfilment of motivational factors is necessary for invocation of higher working performances and their effect on motivation is long-term but its non-fulfilment does not necessarily invoke dissatisfaction (Adair, 2005).

The Alderferøs Theory of Motivational Needs of Existence, Relatedness and Growth (ERG) rearranges the original Maslowøs hierarchy of needs from particular to abstract ones in three grades ó existence (guarantee of employment), relatedness (friendship, relations) and growth (self-fulfilment). In practice, this theory is used for composing or understanding the motivational factors of workers (Cejthamr, 2005).

The McClellandøs Theory of Needs states that people have the need to achieve something, to belong somewhere, and the need of power. The need of performance is defined as a need to succeed in the competition with others, the need of partnership is defined as a need of warm friendly relationships and the need of power is defined as a need to control someone and to influence other people. Each individual has a different level of each need, different need is predominant for each other, however, it does not mean that the other needs are not being demonstrated (Kocianová, 2010).

The Vroomøs Theory of Expectancy states that the motivation is possible only if there is a clearly perceived and applicable relation between the performance and its result, and if the result is considered as a tool for the satisfaction of needs. It explains why the financial motivation works only if there is a direct link between the effort and the reward, in other words, if there is a reward after exerting a certain effort (Tureckiová, 2004).

The theory of equity of J. S. Adams is based on knowledge that the equal treatment means that the individual is treated in the same way as another group of people, where no such group is privileged. The degree of equity is then appreciable by the rate of inputs and returns. The perception of inputs and returns is individual, depending on the values the individuals hold.

Lathamés and Lockeés Theory of Goal Setting claims that the motivation of an individual is higher if he or she is given a particular goal to achieve, if this goal is difficult but manageable, and if there is a performance feedback. According to Locke and Latham the goals influence the performance in four ways: they lead to a higher efforts, they have a positive impact on endurance, they improve focus of attention to the given activity, and they provide motivation for usage of new, not yet well-proven strategies. The individual should participate in the determination of the goal (Armstrong, 2007).

2.2 Research of motivation in the Czech Republic

In the Czech Republic the authors interested in motivation are mainly prof. Milan Nakone ný (2009a, 2009b, 2003, 1998, 1996, 1973), prof. Jan áp and prof. Ji í Mare–(áp, 1997; áp, 1990; áp, Mare–, 2007), prof. Alena Plháková (2004), prof. Marie Vágnerová (2004), all aforementioned from psychological field. Doc. Vladimír Hrabal often together with doc. Isabella Pavelková (Hrabal, Pavelková, 2010; Pavelková, 2005; Pavelková 1990; Hrabal Pavelková, 1989) deal with motivation in pedagogical educational process. The latter one has made several researches. Doc. Zelinková (2007) partly addresses this topic.

Students are influenced by many factors when choosing the field of tertiary study. According to the latest republic research in the Czech Republic from 2005 (Menclová, Ba-tová, 2005, p. 20) the most common motivation for attending a university is a possibility to gain good professional career (61,5% of public universitiesø students). Other factor is a possibility to get well-paid job (61,4%), possibility to gain a university title (52,1%), good position in society (43,1%). Less mentioned motivation factors are a possibility to develop their education, knowledge and skills (34,3%), postponing the leaving into practice and a possibility to live a student life, bigger interest in the specific field or following the family tradition.

3 Methodology

The aim of the thesis is to identify motivation factors influencing motivation of students at the Faculty of Education at the Palacký University in Olomouc for their specialization on teacher profession, and to compare them with motivation of students of teacherøs programmes at other faculties at the Palacký University in Olomouc. Therefore a quantitative research in the form of a questionnaire together with a qualitative method ó interview seems to be a suitable instrument. The questionnaire method will prove useful also thanks to anonymity, because the questionnaire will include sensitive data about the respondents. The particular research tool which will be used for the exercise of this research is not yet available and is currently being developed at this stage. Therefore we are not able to describe it in more detail or to specify the number of items, etc.

3.1 Description of applied method

Chráska (2007, p. 163), Schneider (1974, p. 34), Gavora (2000, p. 99) and Pelikán (2011, p. 104) describe a questionnaire as very frequent method to obtain data in a pedagogical research. This method is often criticised for not focusing on what the respondents are like and for focusing solemnly on how they perceive themselves. We must expect subjective statements (Pelikán, 2011, p. 105). The advantage is its relatively easy application, it is the cheapest among other methods and it easily cover a wide range of respondents (Schneider, 1974, p. 34), other advantage is its easy administration (Pelikán, 2011, p. 105). A questionnaire is õa system of pre-prepared accurately formulated questions which are thoughtfully arranged and the respondent

answers them in written formö (Chráska, 2007, p. 163). When preparing a questionnaire it is essential to divide the problem into several parts, then there are formulated items to every part (Gavora, 2000, p. 99). Items (questions) in the questionnaire may be divided according to several criteria; the most frequent are objective, form of answer and content, which the item concentrates on.

An interview is õa method of data collection on pedagogical reality, which is based on direct verbal communication between a research worker and a respondentö (Chráska, 2007, p. 182). It is one of the methods, which is not based on a written statement of the respondents but on verbal communication (Pelikán, 2011, p. 117). It belongs to the most difficult and the most valuable method for qualitative data collection (Miovský, 2006, p. 155; Schneider, 1974, p. 55). Its advantage compared with other research methods is establishing of personal contact between a researcher and respondents. It is possible to follow respondentøs reactions on the questions and consequently to adjust the interview. Two aspects play crucial role during an interview ó researcherøs approach towards respondent and his attitude to the content of a statement. During an interview it is essential to maintain respondentøs motivation by researcherøs certain manner of behaviour but at the same time the researches must not influence the content of a respondentøs statement. An interview is a method, which is able to obtain facts and to scrutinize respondentøs motives and attitudes.

3.2 The research description

The main aim of dissertation thesis is to identify motivation factors, which influence motivation of students at the Faculty of Education at Palacký University Olomouc for their specialization on teacher profession, and to compare them with motivation of students at other faculties at Palacký University Olomouc.

The other aims of dissertation are:

- 1) To find out which motivational factors influence respondents more often and less
- 2) To find out how many respondents really want to be a teacher
- 3) To find out which motivation factors predominate in students of 1st and 2nd year of the Master's degree program at Faculty of Education
- 4) To find out which motivation factors predominate in students of lifelong education
- 5) To find out differences in the students motivation between faculties
- 6) To find out differences in the motivation between students of Master's degree program and students of lifelong education
- 7) To find out differences in the motivation between students of daily study form and students of combined study form
- 8) To find out which motivation factors predominate in students of 1st and 2nd year of the Master's degree program at all selected faculties (Faculty of Education, Faculty of Science, Faculty of Physical Culture, Faculty of Theology)

The respondents will be students at Palacký University Olomouc, specifically the students of 1st and 2nd year of the Masters' degree programme teaching subjects of both daily and combined study form, and students of teaching subject of Center of Lifelong Education.

We chose Faculty of Education, Faculty of Physical Culture, Faculty of Theology and Faculty of Science. Only these faculties offer the fields of study for teachers.

Students of lifelong education will be from Faculty of Education (there is Center of Lifelong Education), Faculty of Philosophy (there is Lifelong Learning Institute), Faculty of Science (there is Lifelong Learning Programme). Only these three faculties have teaching subject on lifelong education.

As it was already mentioned students are influenced by many factors when choosing the field of study. I expect that by a quantitative research I will find out what motive are the strongest and which are weaker or what portion of students would like to be engaged in teacherøs profession after graduation. For better understanding I present two research questions:

What are motivation factors influencing motivation of students attending 1st and 2nd year of the Masterøs degree programme of daily study form?

What are motivation factors influencing motivation of students attending 1st and 2nd year of the Masterøs degree programme of combined study form?

What are the differences between students attending the Masterøs degree programme of daily and combined study form?

Which motivation factors influence respondents more and which less often?

What portion of respondents wants to perform teacherøs profession in future?

4 Conclusion

There are many factors influencing selection of future profession, they are both internal and external. The question is how many students of teacherøs programmes have chosen the study field because they really want to perform it.

The contribution of this research is in its uniqueness on the national level. The research focused on the motivation to study has been already carried out abroad but according to our observations no research of motivation to study to become a teacher has been carried out in the Czech Republic so far. The research is a supported project of the Students Grant Agency under the number PdF_2012_015 named Motivation of Students at the Palacký University in Olomouc for Their Specialization on Teacher Profession.

5 Resources

Adair, J. (2005) *The Inspirational Leader: How to Motivate, Encourage & Achieve,* London.

Armstrong, M. (2007) ízení lidských zdroj. Praha.

Cejthamr, V. D dina, J. (2005) Management a organiza ní chování. Praha.

áp, J. (1990) Psychologie mnohostranného vývoje lov ka: Celostátní vysoko-kolská p íru ka pro filozofické a pedagogické fakulty. Praha.

áp, J. (1997) Psychologie výchovy a vyu ování. Praha.

áp, J., Mare-, J. (2007) Psychologie pro u itele. Praha.

Chráska, M. (2007) Metody pedagogického výzkumu: základy kvantitativního výzkumu. Praha.

Gavora, P. (2000) Úvod do pedagogického výzkumu. Brno.

Heller, R. (2001) Jak motivovat druhé. Praha.

Hrabal, V., Man, F., Pavelková, I. (1989) Psychologické otázky motivace ve -kole. Praha.

Hrabal, V., Pavelková, I. (2010) Jaký jsem u itel. Praha.

Kociánová, R. (2010) Personální innosti a metody personální práce. Praha.

Lok-a, J., Lok-ová, I. (1999) Pozornost, motivace, relaxace a tvo ivost d tí ve -kole. Praha.

Menclová, L., Ba-tová, J. (2005) Vysoko-kolský student v eské republice v roce 2005. Praha.

Miovský, M. (2006) Kvalitativní p ístup a metody v psychologickém výzkumu. Praha.

Nakone ný, M. (1973) Emoce a motivace. Praha.

Nakone ný, M. (1996) Motivace lidského chování. Praha.

Nakone ný, M. (1998) Základy psychologie. Praha.

Nakone ný, M. (2003) Úvod do psychologie. Praha.

Nakone ný, M. (2009) Psychologie osobnosti. Praha.

Nakone ný, M. (2009) Sociální psychologie. Praha.

Pavelková, I. (1990) Perspektivní orientace jako initel rozvoje osobnosti. Praha.

Pavelková. I. (2005) fiákovská nemotivovanost. In XIII. konference APV "Pedagogický výzkum: Reflexe spole enských pot eb a o ekávání?õ. Faculty of Education, Olomouc, p 217-221.

Pelikán, J (2011). Základy empirického výzkumu pedagogických jev . Praha.

Plháková, A., Nakone ný, M. (2004) U ebnice obecné psychologie. Praha.

Schneider, M. (1974) Úvod do základ sociologického výzkumu. Olomouc.

Sullivan, G., Strode, J. (2010). Motivation through Goal Setting: A Self-Determined Perspective. *Strategies: A Journal for Physical and Sport Educators*, 23(6), 19-23.

Tureckiová, M. (2004) *ízení a rozvoj lidí ve firmách*. Praha: Grada.

Vágnerová, M. (2004) Základy psychologie. Praha.

Zelinková, O. (2007) Pedagogická diagnostika a individuální vzd lávací program. Praha.

Strategic and Doctrinal Framework of Western Balkans Countries

Granit Zela

PhD Candidate, Strategy and Security Department, Defense Academy "Spiro Moisiu" Tirana, Albania, zelagranit401@gmail.com

Abstract. A comparative analyses of the strategic and doctrinal framework of the Western Balkan countries can provide an insight in the general relations between the main stakeholders involved in writing and passing of these documents, their domestic, regional and international security perceptions, understanding level of these documents part of the "Strategic Culture" of the elites and administrations of these states, terminology used, level of adoption, similarities in form and content, deficiencies and achievements that provide ground for the future strategic documents. Elaborating on doctrinal documents in general and more specifically on national security strategies is very indicative and compatible with the analysis in order to see how the countries explain their own role in the security structures in the region: offers us as well an insight on how these states perceive themselves, first and foremost in the regional context. Finally, identifications of the main problems to be addressed in the future contribute to more efficient strategic documents that serve as overarching fundamental documents for the reformation of the security sector and contribution of these states to the security environment of the 21st century

Keywords: Western Balkans, strategic documents, security environment

1 Introduction

Starting from 2000, Western Balkan countries have a prospect of EU membership and after the clear commitment in Thessaloniki Summit in 2003¹, they already were defined as a new geopolitical unit and constitute a Regional Security Sub Complex. The idea of regional security complexes was originally Buzan's, but has been much worked on by Wæver and is now part of the Copenhagen School's collective theoretical approach to security. (Buzan,Wæver 2003).

The Western Balkans is a genuinely unified security-policy region. This all comes from the sole fact that political elites, but also societies that they represent, still do not interpret the fundamentals of the security in a similar manner even on the common, regional level, certainly with different content stemming from one country to another. The study of the national security strategies of Western Balkan Countries embraces this regional approach which provides the means to detect and observe new trends in security sectors in the Western Balkans and their general level of democratization.

The national security policy is a document describing the way state goals are fulfilled either domestically or externally, aiming to provide for state security and through it provide for the individual security of its citizens. (David Law 2007). This dynamic and less normative definition of a National Security Strategy demands first of all a clear and fictional definition of security per se. This can be achieved firstly only through making a distinction between the different levels of security: human security, state security and international security. It demands the articulation and proclamation of a nation-state's aims in accordance with the geographic, economic, social, political, military and cultural resources of a society.

¹According to the EU-Western Balkans Summit - Declaration, "The Heads of State or Government of the member States of the European Union, the acceding and candidate states, Albania, Bosnia and Herzegovina, Croatia, the former Yugoslav Republic of Macedonia, Serbia and Montenegro, as potential candidates agreed to meet periodically at our level, within the framework of a EU-Western Balkan forum, in order to discuss issues of common concern, to review progress of the countries of the region in their road to Europe, and to exchange views on major developments in the EU. In March 2012 EU launched the Kosovo Feasibility Study which follows a decision by the European Union Council of Ministers a month ago which, at the same time, handed Serbia EU candidate status. In this study by Western Balkans we mean Albania, Bosnia and Herzegovina, Croatia, the former Yugoslav Republic of Macedonia, Serbia, Montenegro and Kosovo whilst the analyses of the Kosovo Security Strategy is to become part of this study in the time period of its completion in the near future.

2 Literature Review

In the study, security is given a wide interpretation; the lack of security under this broad definition is considered a situation requiring emergency measures by a government and/or the international community. Factors representing a threat to security are labeled as threats, potential threats, or risks on the basis of their intensity and probability. At the regional level any factor is regarded as a threat if it endangers the existence or normal operation of the majority of the countries in a region.

The achievement of state security should be the initial goal of every national security strategy, which must take into account of all the security dilemmas and the endemic insecurity of the international system. Any concept of state security must also take into consideration the concept and meaning of societal and environmental security as far as the fruition of them cannot be understood without state's supervision and engagement. Therefore would be inevitable (and not at all a theoretical abstraction) a thorough analyses of the international (and regional) system to identify all main conflicting and cooperative trends, to recognize possible opportunities or threats as well as to enter into alliances or avoid threats either from state actors or from non-states ones. But it is also clear that the analyses of security or insecurity descending from the international system cannot be achieved without a clear methodology and without the appropriate epistemological means.

3 Data and Methodology

The methodological use of comparative analyses of comparative analysis of main strategic documents: the National Security Strategy, Defense Strategy and Strategic Defense Review of the countries of the region is justified on several levels: Firstly, the countries of the region had similar experience of instabilities and crises in the last decade of 20th century, as well as a slow stabilization process at the beginning of the new century; Secondly, the timing for the creation of new strategies was again very similar–they have all adopted their first strategic documents in the last 10 years or so; Finally, the countries' answer to the ultimate push from the international community and especially the EU in this sense, for regional cooperation, their sincerity and willingness to cooperate clearly show how far they have moved from the "remnants" of their recent gloomy past. (Hadžić, Timotić and Petrović, 2010).

The study aims at: Studying the security strategies of the Western Balkan Countries in their historical context they were created, the process and the final content of the strategies, the effects and results.

• Contributing in the creation of the expertise in the fields of the strategic thought, discussion of the concepts, principles, the best

international practices of national security documents, terminology, problematic and the innovations they have brought.

- Clarifying understanding of the terminology of the strategic documents, their process of initiation, compilation, execution, assessment (lessons learned) in service of an effective model of security strategy.
- Becoming part of the debate in the community of researchers' by indicating the common security concerns of Western Balkan Countries and distinctions among them regarding the security dilemmas.
- Serving through the publishing of the findings as a way of communication and education of the public on the current and future threats and provide a contribution to the creation of the "strategic culture".
- Reaching conclusion based on the new experience of the Western Balkan Countries with the security strategies.
- Becoming part of the tradition of the Comparative National Security and security studies in the country.

The study is based on a variety of methods selected referring to the resources available; the previous experience as a researcher, the ethical standards and the objective to acquire updated information related to the topic of research. Briefly, these methods include:

- Detailed study of the literature on the security environments of the Western Balkan Countries after the Cold War;
- Study of the critical literature related to our object of study,
- Constant consultation with the PHD thesis advisor.
- Consultation with military and civilian experts of the security field in and out of the country.
- Engagement with a think-tank that has performed studies on this topic of research,
- Presentation of the topics and discussions with the audiences/panels of security issues,
- Participation in Conferences on Security Issues,
- Publishing in academic journals,
- Discussions with professors, scholars and specialists in the security area.

The methodological use of comparative analyses of comparative analysis of main strategic documents: the National Security Strategy, Defense Strategy and Strategic Defense Review of the countries of the region is justified on several levels: Firstly, the countries of the region had similar experience of instabilities and crises in the last decade of 20th century, as well as a slow stabilization process at the beginning of the new century; Secondly, the timing for the creation of new strategies was again very similar–they have all adopted their first strategic documents in the last 10 years or so; Finally, the countries' answer to the ultimate push from the international community and especially the EU in this sense, for regional cooperation, their sincerity and willingness to cooperate clearly show how far they have moved from the "remnants" of their recent gloomy past. (Hadžić, Timotić and Petrović, 2010).

4 Empirical Analyses

The strategic and doctrinal documents of the countries of the Western Balkans have all been created in the first decade of 21st century. After 2000 and until 2010, practically all the countries of the region had their first strategic documents. Albania passed its first strategic document, the Security Strategy Document in 2000, whilst the first document labeled as National Security Strategy was adopted in 2004 and revised in 2007. The National Defense Strategy was adopted in 2000 and revised several times afterwards, the last revision dating to 2007. Finally, the Military Strategy was adopted in 2002 and revised in 2005, while the White Paper on Defense was published in 2005. In Bosnia and Herzegovina, the so-called Defense Policy was adopted in 2001, the Military Doctrine was endorsed in 2003, and the Security Policy of BiH in 2006. The Defense Review was completed in April 2009. Croatian case saw the National Security Strategy and Defense Strategy being passed at the same day in 2002. A year later, the Military Strategy was approved, while the Strategic Defense Review was adopted in 2005.

Montenegro had its first versions of the National Security Strategy and the Defense Strategy adopted only few months after independence in 2006, whilst upgraded versions were adopted two years later, at the end of 2008. Former Republic of Macedonia adopted the Strategy of Defense in 1999 (it was amended in February 2010), then approved the National Security and Defense Concept and Strategic Defense Review in 2003 and the White Paper on Defense in 2005. Finally, the National Security Strategy was adopted in 2008. In this case though, there is another document that needs to be taken into account as part of strategic documents of FYROM – it is the Ohrid Framework Agreement, which was signed in August 2001 and which set a frame that ended the hostilities between the ethnic Albanian minority and Slavic majority in the country. This document "...has included number of provisions on the issue of security and defense...which have guided overall security policies in the country" (Yusufi, 2010).

Serbia, on the other hand, had even bigger delay in adopting its security and doctrinal documents due to political turmoil caused by the self-proclaimed independence of Kosovo and due to different perceptions of security that then main political parties had. This is why it took three years for the National Security Strategy and Defense Strategy to be adopted at the end of 2009. The Strategic Defense Review from 2006 (relevant for the period 2006 -2010) was revised in 2009 as well. (Hadžić, Timotić and Petrović 2006). Finally, the National Security Strategy of Kosovo was approved 2010, as the last security strategy belonging to the first generation of security strategies in Western

Balkans. At this stage of this study, this strategy is not part of our analyses while Albania and Croatia are at this moment developing their new National Security Strategies after the membership in NATO in 2008. It should be noted that Albania is developing at the same time the Defense Strategic Review and is being assisted by US Department of Defense.

4.1 Form Analyses of Strategic Documents

4.1.1 Naming of the Documents: Similarities and Differences

The security documents of the six countries of the Western Balkan region show a great deal of similarity and diversity at the same time. Among the similarities, it may be noted that these documents were adopted quite recently within a short period of time, and that most of them reflect the current security situation of the countries concerned. The number of documents also adds to the impression that the countries of the region have basically created a developed system of security documents. All states possess 1-3 security-related documents encompassing different levels. One can find at least – depending on the wording -3 types of documents of different levels: national security strategies; military or defense strategies; white papers / strategic defense review. (Gyarmati and Stančić 2007).

Starting from their names, the strategic documents of the countries of the Western Balkans do show some similarity. The highest document in the hierarchy of strategic documents in all countries is the National Security Strategy, although this was not the case everywhere at the very beginning (Macedonian National Security and Defense Concept which was replaced with the National Security Strategy). The exception is BiH which has the Security Policy of BiH. Apart from this cover document, all the countries possess another two types of documents: one is military or defense strategies and the other is white papers or strategic defense reviews.

Some of these documents are either revisions of previously adopted documents, or are soon to be replaced by new versions. On the other hand, the similarity of the names of strategic documents (after the inconsistencies that existed at the beginning) could indicate that most of the countries, despite internal differences, have eventually chosen the same or at least similar model of strategic documents to copy and re-shape to their particular needs. This premise is going to be challenged when we start analyzing the structure of the documents.

4.1.2 Drafting of the Strategic Documents

Even the way that strategies and doctrines were drafted and passed, thus creating the strategic system of the Western Balkan countries, is very

indicative of the countries' level of democracy. Firstly, the extent to which the international actors are included into the drafting of these documents is a certain indicator of a country's capability to independently make the important decisions on its own future. An important insight stemming from the analysis of the drafting of these documents is the political elite's attitude towards the media, academic community, and civil society organizations -if the latter are included in the drafting process, or at least if their opinion is heard, we can assume healthy and partner relations between the sides. In contrast to that, the lack of any public discussion and preparation of the documents "behind the closed doors" signifies poor communication and lack of mutual trust between the two (sometimes confronting) sides. (Hadžić, Timotić and Petrović 2010).

Countries of the region, with no exception, had a difficult task in showing their democratic abilities during this process. A striking example for the first step, the drafting of the strategies, would be Croatia. An excellent idea of creating an independent body tasked to write the National Security Strategy which spent a year and a half drafting this document was neglected in a matter of days and a small group of people from several state ministries wrote the Strategy within few weeks.

Serbian case shows that there is a need for a wide consensus at least among the elite and proper procedures in order for a strategy to be written and adopted. During the period 2006-2007 there were two teams, two Working Groups, one of the President and one of the Prime Minister which produced as many proposals. The proposal of the Presidency's cabinet and the Government's one reflected somewhat different and even (in certain cases) contradicting ideas of the two main political actors in Serbia at that time on the important strategic decisions related to the approaching self-proclamation of Kosovo's independence and other security issues. It was only when it was clearly stated that the Ministry of Defense was in charge of drafting the document and the National Parliament of adopting it, that this problem was sorted out in a proper way, with significant and a hardly acceptable delay, though. (Hadžić, Timotić and Petrović 2010).

4.1.3 Hierarchy of Adoption of the Strategic Documents

The adoption of strategic and doctrinal documents in the right order shows the maturity of political elite and their understanding of the importance of encompassing a strategic doctrinal framework in a proper way. At the top end and first to be developed should be the National Security Strategy and all the other documents should be subordinated to and in line with it. The necessity of adopting this overarching strategy first lies in the fact that it would give the framework for all the others. In this sense, politicians of the Western Balkans countries have shown that, at least at the beginning of this process, they just wanted to gain democratic "credentials" by the mere adoption of strategies, so they were pushing to finish the process as quickly as possible without paying attention to the right order and to the fact that the strategies need to follow this

basic hierarchy. This is why most of the strategic documents do overlap in their content in the manner that the same expressions and wordings are used, which is not that problematic, or simply by copying the content (especially challenges, risks and threats) which is much worse. (Hadžić, Timotić and Petrović 2010).

None of the observed countries followed the right order of drafting and adoption, i.e. the lack of transition from general documents to more specific ones is a rule. For instance, Croatia adopted its first National Security Strategy and Defense Strategy at the same day, 19th March 2002. Serbia did the same with its own documents of the same importance, adopting them in October 2009. Again, this resulted in certain overlapping – some segments in the Defense Strategy have not been elaborated enough, but entirely "imported" from the National Security Strategy (NSS).

Further on, the Serbian Strategic Defense Review was adopted 7 months prior to these two documents, contributing to somewhat chaotic situation. Montenegro, albeit being among the last ones to adopt strategic documents, has already "had a chance" to make the same mistake, and afterwards to avoid repeating it. The NSS and Defense strategy were firstly adopted within 20 days in 2006 – the NSS on November 27th and the Defense Strategy on December 17th, not giving enough time for the latter to be made in line with the NSS. Even more confusing situation occurred in Bosnia though, where the first strategic and doctrinal documents defining BiH's strategic security goals were the Defense Policy and Military Doctrine. Nearly three years later, the Security Policy (the most general security document, counterpart of the NSS) was adopted, although logic presumes the reverse order. It is only in Kosovo where the current situation justifies the so-called "bottom-up approach", meaning drafting and adoption of sector-based strategies instead of development of a national security strategy.

Former Republic of Macedonia seems to make the same case as all other countries. Not only the first Defense Strategy and White Paper on Defense were adopted before the NSS (or National and Security Concept, as the first version of the NSS was called), but all other documents (apart from Ohrid Framework Agreement, whose adoption was dictated by the conflict and efforts for its resolution) were adopted in the same year, 2003 (with White Paper on Defense and Strategy for the Police Reforms amended in 2005 and 2004, respectively), although not at the same time. Such a tight schedule does not leave a lot of faith in the capabilities of the authors to make a sound coordination and right hierarchy. Still, despite these disturbances, in principle the strategic framework of Former Republic of Macedonia is by far the most harmonized in the region. For instance, the Strategic Defense Review and the revised White Paper on Defense explicitly state their subordination to the National Security and Defense Concept and explain that they follow the views, positions and guidance set out in it. On the other hand, "the Defense Strategy refers mainly to the National Security Strategy" (Yusufi, 2010).

4.2 Content Analyses of the Security Strategies

The various security documents of the countries of the region are hardly comparable, partly because of their different genre and consequently their size and subject matter, and partly because they are structured diversely. A comparison of the content structure of these documents shows the complete lack of any model upon which these strategies were based. A positive interpretation of this situation could be that these new actors of international relations have formulated their strategies in a truly independent manner. However, from possibly a more realistic perspective it could also be considered quite problematic that countries aiming to accede to the same international security organizations have not elaborated more consistent and more similar security documents. In addition to the chaotic nature of the content structure, it can also be demonstrated that there are important elements missing in the general outlines of certain documents. As to the principles, values and interests, goals, and security environment situation, most states (four out of the six) explicitly deal with these issues, but only half of them dedicate a specific chapter to the most important element of threat perception i.e. challenges, risks and threats. (Gyarmati and Stančić 2007).

4.2.1 Assessment of the Security Environment in Security Strategies

Using modern terminology, the security documents of the Western Balkan countries generally describe the security environment in a similar way, interpreting post-bipolar era international security as a complex phenomenon (referred to as multidimensional in the Croatian defense strategy document). The security situation is also characterized as changing in a dynamic and often unpredictable way. NATO and the EU are unanimously regarded as basic factors of stability, even if countries are at different stages in their efforts to join these entities. Emphasis is placed on the commonality of values and objectives of the countries of the region, and on the need and desire for regional cooperation.

The analytical framework of the current security environment also shows similarities. The Albanian National Security Strategy defines historical legacy, geostrategic importance and ethnic fragmentation as key factors in interpreting the current security situation. Most security strategies underline that incomplete transition processes also represent a key potential source of instability. n this regard, the process of reconciliation is described as unfinished business in the Macedonian document, while the relevant Albanian document states that confidence-building measures play a crucial role in furthering reconciliation. The importance and special security role of the Western Balkan region is defined in a partly similar and partly different way in the relevant documents. The Albanian document refers to the region's "bridge role" as an important factor in this respect. The Bosnian paper describes the country's position as one "connecting the Danube region with the Mid-

Adriatic region", while the Macedonian National Security and Defense Concept labels its immediate environment – in a somewhat exaggerated manner – the "crossroad" between Europe, Asia and Africa. The Croatian document identifies the country as occupying a "multidimensional regional position" connected simultaneously to Eastern Europe, South-Eastern Europe and the Mediterranean regions.

Although there are different wordings in the documents as to the current security situation in the Western Balkans, they all characterize the situation in a similar way. The Croatian Security Strategy states that South-Eastern Europe is the "most troubled corner" of the continent. Similarly, the Macedonian National Security and Defense Concept acknowledge that South-Eastern Europe "remains the most unstable region in Europe", while the Serbian Strategic Defense Review describes the region as the "least stable" in Europe, where the possibility of the use of force has not disappeared completely. Certain documents make specific inputs to the overall security picture. The Macedonian document is unique in mentioning the "hostile role of foreign security services". The Albanian security strategy, for example, introduces the "Albanian factor" as an important security element. The Croatian security strategy describes some of its neighbors in terms of security, and refers to the situation in Bosnia and Herzegovina as "delicate", and notes that although Yugoslavia has been the "main cause of regional instability over the last ten years, it is a "potential partner".

4.2.2 Threat Perception in Security Strategies

The national security documents examine the perceived threats, challenges, and risks faced by their countries. In addition to these quantitative problems, there also seems to be a qualitative problem reflected in the markedly diverse way these documents classify - if they do so at all -threats, challenges and risks. Two countries (Albania and Bosnia and Herzegovina), and to a certain extent Croatia, classify the challenges, threats and risks in a structured way, using three levels of analysis: local, regional, and global. The Croatian security strategy refers to the past ten years, which is the reason for mentioning of "Yugoslavia". Two other countries (Macedonia and Serbia) resort to another method of classification, using the level of intensity indicator, combined with a chronological approach (short-, mid- and long-term), with an additional military–non military distinction in the Serbian case.

The threat assessment section of Montenegro's security strategy – relatively weak even by Western Balkan standards – makes no attempt to classify threats and challenges. The Montenegrin documents also fail to interpret general threats in the Montenegrin context, and the way they are formulated could be applied to any country in the world because they do not contain any specific reference to the Republic of Montenegro. Another aspect of the qualitative problem with regard to these basic security documents is a lack of prioritization when listing threats and challenges, as well as the failure to

differentiate between threats, risks, and challenges. When comparing the threat perception of the documents we will rely on the classification used by the best structured documents, applying the internal-regional-global categorization to the other documents. (István Gyarmati and Darko Stančić 2007).

4.2.3 Risks and Threats presents in Security Strategies

An observation stemming from the reading of the documents is that none of them makes a clear and overall distinction between challenges, risks and threats. This makes a proper risk analysis of various documents even more complicated, leaving the researchers to make their own conclusions and interpretations from the reading and general ideas of the texts. For instance, Serbian NSS clearly states that "non-legal, unilaterally declared independence of Kosovo represents the biggest threat to the security of the Republic of Serbia" (NSS of Serbia, 2009), whilst the other enumerated challenges, risks or threats are not always clearly put into one of the categories. The negative side of this lack of lucidity is, of course, not due to the problems they raise for researchers to read them, but in the fact that they do not fulfill their primary purpose then – the main stakeholders in security sector do not get a clear vision on how to react to a certain problem and how to be prepared to overcome it in advance.

Another important part of the analysis that stands for most of the countries is the lack of hierarchy between challenges, risks and threats enumerated in the documents. This serious problem means that again the readers of the documents cannot anticipate immediately what is perceived as the biggest threat for the country. For some of the documents, we can assume that the order of presentation in the strategies actually represents a certain level of prioritization, but we cannot be certain about it. Other documents do state what the biggest threat to their country is, but leave the other CRTs out of any hierarchy. This is the case with Serbia: "non-legal, unilaterally declared independence of Kosovo represents the biggest threat to the security of the Republic of Serbia". Then again, unless this threat is put as first on the list, which is not the case in Serbian NSS (it is on the third place), we can then exclude the possibility to perceive the very order of presentation as a sort of hierarchy. Croatian NSS makes the problem even worse by dividing the enumeration of the CRTs in two chapters. The first chapter mentions some of the CRTs, whilst they are being enumerated and rephrased in the third chapter. The only presumption that can be made is that those CRTs mentioned in the first chapter have a higher "ranking" in the prioritization than the others. Still, this does not fully solve the problem of the necessary prioritization, because even among the two groups of CRTs there is again no differentiation by importance. (Hadžić, Timotić and Petrović 2010).

The only country where some sort of hierarchy of perceived CRTs is the Former Yugoslav Republic of Macedonia, where they are put in time

dimension (currently, mid-term, long-term) and sorted by the level of intensity (high, medium, low, very low). In this hierarchical structure, the top place (the most significant CRT) goes for "possible manifestations of extreme nationalism, racial and religious intolerance, international terrorism, organized crime, illegal migration, and illegal trade with all types including trade with strategic and dual use of materials, insufficiently secure and efficient borders etc." This prioritization exists only in Macedonia's Strategic Defense Review of 2003. Since none of the other documents repeats this prioritization, thus not making coherent overall strategic framework, it loses a lot from its relevance.

Having noticed that, the Albanian think-tank "Institute for Democracy and Mediation" has prepared a Policy Paper to assist in the process of the review of the National Security Strategy in Albania that commenced in December 2011 and is in the process the CRT analyses of the security environment. Conclusions will be made after this process ends, but it is to be noted that unfortunately the process of drafting the National Security Strategy 2012 and The Defense Review are taking place simultaneously which leads us think that still the "Strategic Culture" is not consolidated and lessons learned are now drawn.

4.2.4 Internal Threats depicted in Security Strategies

Two CRTs emerging at the national level for nearly all the countries are disasters and transitional problems, whilst other two of them, organized crime and terrorism exist as perceived internal threats in all the countries except BiH and Montenegro. The creators of Montenegrin strategic documents might have intentionally done that in order to "avoid" emphasizing organized crime as an internal threat; it is unclear why it is not a part of internal threats in Bosnian document. The usage of terrorism as an internal threat is somewhat ambiguous - it is very hard to prove the existence of terrorist "cells" in the region and since the conflicts in the southern part of the Western Balkans - Kosovo, southern Serbia and Former Yugoslav Republic of Macedonia have ended, there were no real terrorist threats in the region. Transitional problems are thoroughly described in the documents as problems of "political transition which result in a slow development of effective and efficient executive, legislative and judicial authorities; problems of the transition to market economy, which result in a low level of domestic and foreign investment and are favorable for grey economy and black market; slow pace, difficulties and irregularities in implementation of privatization process...technological regression and deterioration of production capacities...unemployment..." (Bosnian Defense Review - not yet published). Document of Former Yugoslav Republic of Macedonia add to this list some peculiar threats - "urban terrorism, serious crime including blackmail, racketeering, murders and attacks on the property of citizens, economic crime, tax evasion..."(The National Security and Defense Concept, Macedonia). Inclusion of disasters (ecological, technological and epidemics) as internal threats shows that the countries of the region do follow at least a minimum of standards regarding contemporary security issues.

Bosnia defines several threats that are direct consequences of the war held on its soil, some of them in a very interesting way: firstly, there is an "incomplete and selective implementation of the Dayton Peace Accord". The wording here is probably the lowest common denominator that was agreed upon, since the two confronted sides have firm, stubborn and completely opposite stances on the issue. Finally, the authors of Albanian and Serbian NSS have paid special attention to demography problems, i.e. illegal migration that leads to "brain drain" effect, as well as "uncontrolled population movement" (Albanian NSS), a phenomenon present in all Western Balkan countries, but specifically emphasized only in these two strategies. The Serbian NSS had just a bit different wording, stating that the problem is the "non-proportional economic and demographic development of the RS and neighboring countries" leading to migrations from underdeveloped regions to the more developed ones (NSS Serbia, 2009). (Hadžić, Timotić and Petrović 2010).

4.2.5 Regional Threats depicted in Security Strategies

When reflecting on regional threats, the countries of the region show the highest degree of similarity in their documents. This comes from the notion that the legacy of the wars for socialist Yugoslavia's heritage and historical factors are still present and constitute a significant burden for creating a favorable security environment, which is acknowledged throughout the region. Therefore, the documents in different variations state that, although the possibility of an armed Conflict is reduced, "...it can never be entirely ruled out" (NSS of Montenegro). The degree of certainty that there is no possibility for another regional "chaos" does differ though – Albania is neutral in this sense, whilst only Macedonian documents speak of the "realistic risks and dangers" coming from the regional "national, religious, greater-state and territorial confrontation". BiH finally, in the scope of its own fears, emphasizes the "aspiration for secession, autonomy and independence of certain ethnic groups" and armed conflicts that can arise from these aspirations as a serious regional (let alone national) problem.

A careful reading of the documents shows that the articulation of regional threats stems from internal problems that the countries have. Naturally, a lot of these problems have their roots in the fact that the region is still labeled as a post conflict area, burdened with serious social, political and economic problems. On the other hand, the perception of regional threats gives us, even between the lines, the picture of very complicated relations among different countries in the region. Unfortunately, first neighbors are still in most cases perceived as "the others" that are not an imminent threat, but might be a destabilizing factor in the (near) future.

Due to this fact and a lot of other unresolved issues, the regional instability and crises are one of common regional threats for all the countries. It seems that, although peace, some sort of fragile stability and technical democracy have been achieved, even the countries themselves (let alone the EU and the rest of the international community) are not yet certain if all the problems and troubles are way behind, or likely to happen again. Specific in this sense is the case of BiH and Serbia which emphasizes "secessionist tendencies", each of the two burdened with their own problems. On the other hand, most of the countries refer to organized crime and extremism as regional problems. Former Yugoslav Republic of Macedonia has as a regional threat a "possible manifestation of extreme nationalism, racial and religious intolerance", while Serbia states that "national, religious and political extremism, and destruction of cultural heritage" characterize the state of affairs of security in the region, thus "burdening the process of democratic transition" in its countries. Transition is also being mapped as the regional and not just an isolated country problem. Actually, the strategies stipulate that transition causes various political, social and economic cleavages that, they themselves can be source of regional instability. Apart from the regional threats that are being put in the documents by all the countries, other (important) issues are also being raised by only some of them. Among the most interesting is certainly the issue of energy routes and a possible instability if energy imports are stopped. This is a threat on both national and regional levels as perceived in the Macedonian National and Security Defense Concept.

The last but not the least important, there are issues that have been tackled very briefly and without a lot of explanations, but that do complete the overall picture of how regional threats are perceived from a country's perspective. Serbia, for instance, elaborates a bit on a tough position of the refugees and IDPs and their unresolved status in the region. On the other hand, BiH and Croatia rightfully point to a huge number of anti-personnel mines and unexploded ordinance on their territory and the territories of the neighboring countries. Generally speaking, this part of the CRT analysis tends to be the most important one – if the countries are capable of understanding what the main obstacles on the regional level are, thus showing at least a certain level of maturity and mutual understanding, their Euro-Atlantic (or European, in case of Serbia) prospects are much brighter.

4.2.6 Global Threats depicted in Security Strategies

This is the least elaborated category of all. The world's "usual suspect" in the last nine years –terrorism - is elaborated on in every strategy as the biggest threat of the modern humanity, with possible effects for the Balkans as well. The same goes for organized crime, which (as we have already stated) works as well in the Balkans as in the other parts of the world. Every document does also mention the possibility of interstate armed conflicts, but emphasizes that there is a low probability for this to happen. The Macedonian Defense Review even gives an estimation that, in the long-term (10 years and beyond), this

probability will further decrease, as well as the probability of the "nonconventional and asymmetric threats, risks and dangers". Still, the Review does not explain the basis for this estimation and especially the source of optimism for the latter, since none of the indicators points to a decrease in the possibility of asymmetric threats.

Environmental challenges are mentioned in all the countries as a problem that can have a serious effect on the security. Among the most mentioned environmental problems are: pollution, degradation, climate change, but also the scarcity of natural resources and the potential for future clashes arising from the possibility of having the monopoly over their use. The enumeration of global threats in the regional strategic and doctrinal documents clearly shows signs of copying from other documents of that kind, or using documents of other countries as models. As we have already mentioned, this is not strange for the region – the countries that have become part of NATO are either using this organization's documents to show that they are in line with its perception of the CRT, or waiting for new strategic documents to be adopted to do the same. On the other hand, several countries in the region were (or still are) to some extent the protectorates of the international community, meaning that the first versions of their strategic documents were written by the members of the international community and simply adopted by (selected) local stakeholders. Finally, even the countries that don't fall into the two mentioned categories (vet), do try to comply with the ideas and standards of the Euro-Atlantic community, perceiving it as their natural surroundings.

4.2.7 Perceptions of Regional Security Cooperation in Security Strategies

The security strategy documents devote varying levels of attention to the issue of regional security cooperation. In Albania, the so-called "Albanian national issue" is a high priority on the list of security strategy objectives, and the relevant Albanian document suggests that "the Albanian national issue will be achieved through European and Euro-Atlantic integration of the countries of the region". Other than this, Albania's national security strategy papers do not deal with regional security cooperation.

Likewise, Bosnia does not elaborate specifically on regional cooperation in its strategic documents. Only one paragraph and minor references shed some light on its regional cooperation approach. In the section of the document dealing with defense policy it is stated that the country is "committed to promote its role and importance in regional cooperation", especially regarding implementation of the Agreement on Regional Arms Control. In another section dealing with the principles of defense policy, one finds reference to the "balance of forces and capabilities in...South East Europe". By contrast, Croatia's national security strategy devotes appropriate attention to and explains the goals of security policy at regional level. The regional level is first mentioned among its national security principles in the context of security threats originating from that level. In the sub-chapter dealing with neighborly

and regional relations, Croatia's relevant strategic document then states that regional cooperation "is an important component of Croatian foreign policy". Here, once again, emphasis is placed on Croatia's self-perceived "multiregional identity".

After enumerating the ongoing regional initiatives, the document expresses interest in "deepening all forms of cooperation" in the region. With regard to international peacekeeping missions, it should be noted that Croatia's readiness to "participate in common military units" is referred to exclusively with regard to the "Northern tier" of its self-proclaimed multiregional identity, namely the Central European states, thus apparently excluding the potential for a joint Balkan effort. The Macedonian National Security and Defense Concept is rather vague on regional cooperation (on other issues as well). It refers to this issue in a declarative manner, expressing interest in "deepening and finding new forms of cooperation".

One section is more specific, and stresses the need to "care for the protection and the permanent improvement of the freedoms and rights of the Macedonian minority living in the neighboring countries". The Montenegrin Strategy of National Security gives the impression that this country is more directly a part of the world than of its immediate neighborhood. In this short document, the regional level is mentioned twice in the context of possible threats generated by transition processes. Other than this, it makes no attempt whatsoever to elaborate on regional cooperation. (Hadžić, Timotić and Petrović 2010).The Serbian Strategic Defense Review devotes two paragraphs to the issue, stating that the "country takes (an) active part in regional initiatives and programmes". The document also expresses the view that "joint crisis management and the establishment of regional mechanisms for the risks and threat prevention are significant prerequisites for security consolidation".

5 Conclusions

Over the past half decade the Western Balkan countries have elaborated and developed an impressive number of security strategy documents. In most cases, these documents constitute a multi-level system of strategic approaches reflecting the three standard levels of such papers, namely national security strategies, military/defense strategies and white papers/defense reviews. Furthermore, these documents have constantly been revised and updated. Naturally, the scope of our analysis needed to be narrowed down only to the main strategic security documents. The comparison of the types of documents, the content structure, the topics dealt with, and the level of adoption leaves you with the overall impression that dissimilarity and variety prevails over similarity. More importantly, it reveals that these documents did not follow any unified model. When assessing the security environment the national documents typically make two points that underline the strategic importance of the Western Balkan region.

Firstly, there is, at least in the eyes of an outside observer, a somewhat exaggerated self-perception that the region is an extremely important "crossroad" between continents, and plays a "bridge role", giving the impression that the "region in transition" is in reality a "region of transit". Secondly, and more realistically, the documents state unanimously that the Western Balkans constitute the most problematic and unstable region in Europe. As to the core issue, i.e. threat perception, and in spite of the modern terminology and the great length of the texts that characterize certain documents, their level of elaboration on this issue is somewhat limited.

The main problems in this respect are: lack of differentiation between challenges, threats and risks; lack of classification of the challenges, threats and risks; lack of prioritization of the threats; lack of distinction between internal, regional and global threat levels; lack of interpretation of new threats such as terrorism in the local context; mixing up basic notions - e.g. collective security vs. collective defense. When one compares the respective documents of the six countries of the region and their analysis of internal, regional, and global security threats, a progressively greater convergence of view on threat perception appears to emerge as one move towards the broader scenario. In other words, we find the most diverse picture at the internal level, which becomes more homogenous at the regional level, and almost identical at the global level.

References

Law D, (2007) Best practices in the field of national security strategy (eds) Toci F, Mbi zhvillimin e Strategjisë së Sigurisë Kombëtare, Institute for Democracy and Mediation, Tirana.

Gyarmati, I. and Stanèiæ, D. (2007), *Study on the Assessment of Regional Security Threats and Challenges in the Western Balkans*, Geneva Centre for the Democratic Control of Armed Forces (DCAF), Geneva.

Hide, E. and Camberi, H. (2010), *Albania*, (eds). Hadžić, M, Timotić M and Petrović P, *Security Policies in the Western Balkans*: Centre for Civil-Military Relations, Belgrade.

Dautović K, (2010), *Bosnia and Herzegovina*, (eds) Hadžić, M, Timotić M and Petrović P, *Security Policies in the Western Balkans*: Centre for Civil-Military Relations, Belgrade.

Mahečić Z, (2010), *Croatia*, (eds). Hadžić, M, Timotić M and Petrović P, *Security Policies in the Western Balkans*: Centre for Civil-Military Relations, Belgrade.

Qehaja, F. (2010), *Kosovo*, (eds) Hadžić, M, Timotić M and Petrović P, *Security Policies in the Western Balkans*: Centre for Civil-Military Relations, Belgrade.

Yusufi, I. (2010), *Macedonia*, (eds) Hadžić, M, Timotić M and Petrović P, Security Policies in the Western Balkans: Centre for Civil-Military Relations, Belgrade.

Radević R, (2010) *Montengero* (eds). (eds) Hadžić, M, Timotić M and Petrović P, *Security Policies in the Western Balkans*: Centre for Civil-Military Relations, Belgrade.

Bjeloš M, (2010) Serbia, (eds) Hadžić, M, Timotić M and Petrović P, Security Policies in the Western Balkans: Centre for Civil-Military Relations, Belgrade.

Abusara A, (2010) *Comparative analyses*, (eds) Hadžić, M, Timotić M and Petrović P, *Security Policies in the Western Balkans*: Centre for Civil-Military Relations, Belgrade.

Ferenc Gazdag, László Póti, Judit Takács, and Péter Tálas, (2007) Assessment of the Security of the Western Balkans and a Comparative Analysis of the Threat Perception in the Countries of the Region" (eds) (eds) Gyarmati I, and Stančić D, Study on the Assessment of Regional Security Threats and Challenges in the Western Balkans, Geneva Centre for Democratic Control of Armed Forces, Geneva.

Sotiraq Hroni and Gen.Lt (ret) Pellumb Qazimi, View from Albania (eds) Gyarmati I, and Stančić D, Study on the Assessment of Regional Security Threats and Challenges

in the Western Balkans, Geneva Centre for Democratic Control of Armed Forces, Geneva.

Olga Palinkasev, *View from Bosnia and Herzegovina*, (eds) (eds) Gyarmati I, and Stančić D, *Study on the Assessment of Regional Security Threats and Challenges in the Western Balkans*, Geneva Centre for Democratic Control of Armed Forces, Geneva.

Mladen Staničić, Ph.D. *View from Croatia* (eds) Gyarmati I, and Stančić D, *Study on the Assessment of Regional Security Threats and Challenges in the Western Balkans*, Geneva Centre for Democratic Control of Armed Forces, Geneva.

Aleksandar Matovski *View from Macedonia*, (eds) Gyarmati I, and Stančić D, *Study* on the Assessment of Regional Security Threats and Challenges in the Western Balkans, Geneva Centre for Democratic Control of Armed Forces, Geneva.

Stojanović S, *View from Serbia*, (eds) Gyarmati I, and Stančić D, *Study on the Assessment of Regional Security Threats and Challenges in the Western Balkans*, Geneva Centre for Democratic Control of Armed Forces, Geneva.

Rearranging the puzzles of security in the Western Balkans: a comparative study of intra-state armed groups

Armend Bekaj

South-East European Research Centre (SEERC), International Faculty, The University of Sheffield armend.bekaj@gmail.com

Abstract. This research aims to examine the post-war rearrangement of security mechanisms in the Western Balkans. specifically in Kosovo, the former Yugoslav Republic of Macedonia (FYROM) and Serbia. As a way of prelude, the research will delve into a brief historical context, so as to provide an insightful analysis on the power dynamics and the oppressive nature of the Yugoslav state in the 1990s, as well as the precarious inter-ethnic balance from the end of the 1990s onwards. By employing a qualitative and comparative analysis, the research will seek to address the challenges and opportunities of reintegration of intra-state armed groups in these three countries, and their role in democratization and security reform processes. Participatory observation, field research and semi-structured interviews will provide for a substantial part of the first-source research. The findings will aim to shed light on the basic interplay between state actors – government, military, police, intelligence services – and the socalled non-state actors – in this case the military and guerrilla groupings which emerged in Kosovo, FYROM and Serbia in the second half of '90s and the first half of '00. Specifically, the research will critically examine how the latter actors, turned into politicians, have become factors for change in the region. The emphasis will be placed on the nature of that change, or in examining the transitory societies of the above-noted Western Balkan countries, specifically democratization processes and security sector reform.

Keywords: Kosovo, Serbia, FYROM, security, democratization

1 Introduction

The Yugoslav wars of the 1990s were accompanied by the emergence of various armed factions that claimed to fight in the name of one ethnic group/nation or another. The Yugoslav Army was monopolized by Belgrade, who considered itself the direct descendant of Yugoslavia. The security vacuum that was emerging with Yugoslavia's disintegration, coupled with the intensification of hostilities between the former Yugoslav entities in which Serbia possessed a clear military advantage, precipitated the creation of national armies in Slovenia, Croatia and Bosnia-Herzegovina. In the mid-1990s, the same pattern was unfolding in Kosovo: the emergence of the Kosovo Liberation Army (KLA) was a manifestation of the Kosovar Albanians' reaction against the oppression by the Serbian regime (Clark 2000; Ramet 2000; Glenny 2000; Little & Silber 1995). In fact, the emergence of this armed resistance was only a different and, in many respects, a complimentary manifestation of a civic and nonviolent resistance which had come to life in the early 1990s. However, the perceived failure of the nonviolent resistance to reach its objectives, contributed in part to the creation of the KLA.

The intensification of the armed conflict led to an increased interest in international diplomacy, first in Croatia and Bosnia-Herzegovina, and then a few years later in Kosovo. Eventually, international diplomacy was followed by NATO's armed intervention in Serbia. The end of the Kosovo war brought to an end the violent cycle of the break-up of Yugoslavia, resulting in new and reconfigured frontiers in the Western Balkans.

However, at the outset of the twenty-first century we saw the appearance of two additional ethnic Albanian intra-state armed groups in newly-constituted states: one in the Former Yugoslav Republic of Macedonia (FYROM) (National Liberation Army – NLA) and another one in the southern part of Serbia (Ushtria Çlirimtare e Preshevës, Medvegjës dhe Bujanocit - UÇPMB). Short-lived conflicts between them and the national armies of the two respective countries ensued, with the former claiming to fight for rights that were suppressed, and the latter claiming to fight for their community's rights and exert control and stability over their territories. In the end, similar to the KLA, the NLA and UÇPMB were involved in peace processes, with some of their leaders eventually becoming political figures.

At present, FYROM continues to face obstacles to EU and NATO integration, largely due to its long-standing impasse on the name issue with the neighbouring Greece. Moreover, although the country is striving towards building a truly interethnic society with its Albanian and other ethnic minorities, there is some way to go to implementing fully the Ohrid Agreement on decentralisation and local selfgovernment reforms (Ringdal et al, 2007). On the other hand, the conflict between Serbia and Kosovo is in many ways the longest and the most perseverant one. Four years after the declaration, Kosovo's independence is still an ongoing political process. Serbia's adamant stance not to recognize the 'break-away province' and its tireless lobbying against independence, has clearly slowed down the wave of recognitions, and blocked Kosovo's membership accession to certain international bodies, such as the United Nations, the European Union, the Council of Europe, or the Organization for Security and Cooperation in Europe.

Internally, Serbia, FYROM and Kosovo face the common challenges of addressing the reintegration of former combatants of intra-state armed groups. The former members of the armed group UÇMPB in Serbia remain largely un-integrated, while some of them are prosecuted for alleged crimes committed during the conflict. FYROM has gone a long way to reaching a joint agreement with the armed group NLA, whose leaders have been converted into political figures. However, there are lingering accusations of selective justice being exercised upon the former combatants, in spite of an agreed state amnesty. In Kosovo, the transformation of the KLA saw many of its key figures turn politicians or join post-war security mechanisms. Many of them were recruited into the former Kosovo Protection Corps (KPC), and more recently into the Kosovo Security Force (KSF), while others joined the Kosovo Police or private security companies.

This thesis attempts to identify the common denominators between the three armed groups:

- They represented the ethnic Albanian population in each instance;
- They were inter-linked by ideology, manpower and strategy;
- Their representatives occupy space in the current political arena at national or municipal level, etc.

The factions of organised crime and underworld activities are also a spill-over effect of their demobilisation and deviant reintegration, with continuing regional ramifications (Berdal & Malone 2000; Clewlow 2010). We identify this as a *security paradox* within international relations of transitional societies. While these societies are expected to abide by democratic and pluralistic principles of governance, they are simultaneously faced with the daily challenges of integration of former combatants in a meritocratic system, which system might often require bypassing those same persons who fought in wars. Tackling issues of corruption and organized crime are intrinsic and corrosive side-effects of the transition (Thachuk 2005; Tanner 2007).

2 Literature review

The above overview is grounded on literature (Weller 1999; Ramet 2010; Djokić & Ker-Lindsay 2011 etc.), as well as on journalistic and public writings of the late. However, the research proposed here is addressing a yet-unexplored area, which is the interaction of the three intra-state armed groups in Kosovo, Serbia and FYROM, their representatives' transformation into politicians of today, the implications of crime in

politics and public life, and their role in bringing about, or hampering, sustainable peace in the Balkan region - and by default, further integration in Europe. By consequence, the research aims to address *the complexity of securitisation in modern/transitory societies*, and contribute knowledge relevant to disciplines of politics, international relations, security, criminology and social policy. Today, the issue of reintegration of former combatants, their transformation into political decision-makers, or the spill-over consequence into crime, are prevalent for international relations and academia.

3 Data and methodology

My tentative literature base will be derived from eclectic multi-disciplinary sources in human security (Zedner 2003; Kaldor 2007; Dudouet 2009), anthropology (Nordstrom 2007, Ignatieff 1993) and conflict resolution (Miall et al 2003; Lederach 2002).

This research triangulation will be placed at the intersection between security and conflict resolution, by taking an ethnographic approach and applying:

Desk research and literature review – I will be looking for first-hand information from local authors in Kosovo, Serbia and FYROM. My mother tongue is Albanian. I speak fluent Serbian and am conversant in Macedonian. This will enable me to make use of a wide array of valuable texts in local languages. Additionally, English literature will provide a backbone to the desk research. Moreover, I will expand my research to include texts from politics, history, normative and legal acts.

Field research in Kosovo, Serbia and FYROM – this first-hand experience will enable me to conduct a mapping of all security-related actors in the region and their cross-border relations. While in the field, I will gather substantial quantitative data; however, since the nature of my research is more qualitative-driven, I will particularly focus on the latter.

Participatory observation – while spending time in the field, almost daily interactions will occur with decision-makers and other actors who are currently responsible for overseeing security mechanisms in their respective countries, and their possible cross-border cooperation. I will observe from up close the changes introduced in the security-related legal frameworks, and assess the impact of those changes in the ground.

Semi-structured interviews – I will engage with actors involved in armed groups and politics respectively in Kosovo, Serbia and FYROM. By stressing the aspect of narrative interviewing, the interviewees will be encouraged to tell their stories during conflict and the subsequent integration in politics and public life. I expect to meet with representatives from the field of politics, armies, police, and other defence and security bodies, civil society and media. I anticipate conducting approximately 15 interviews in each country. Mindful of the potential ethical issues related to this work, throughout my research I will place careful emphasis on applying the following tools:

Trust and Confidentiality – a form will be signed between the interviewee and interviewer to attest respect for confidentiality.

Anonymity and consent to participate – the rules of interviewing and disclosure will be agreed with interviewees prior to engaging with them.

Knowledge of criminal and other illegal activities - being aware of groups operating illegal activities in the region, I will be cautious in interviewing and obtaining data only from credible and reliable sources.

Personal safety - I will always try not to place myself in unpleasant situations, whenever interacting with people in the ground. I will also make sure to walk away from any potentially threatening situations that I might encounter.

Objective assessment – I am mindful of the fact that my thesis might generate divisive and heated debates, depending on whose 'side' one is perceived to be. In accordance with my academic principles of serving to the truth, I will employ a rigid method of deconstructing narratives and assembling my arguments based on objectivity and empirical evidence. I will seek out the advice of my supervisors to maintain a balanced and unbiased view throughout my academic research.

4 Empirical analysis

My hypothesis will suggest that a sustainable security framework and full-fledged democracy in Kosovo, Serbia and FYROM result directly from the full and nondiscriminatory inclusion of former combatants into the implementation of peace processes. But while the respective governments are called upon to be more open and fairer to the former representatives of intra-state armed groups turned politicians, the latter are expected to make a clear differentiation between politics and crime elements in their midst. Thus, a fairer and more inclusive collaboration between governments and former combatants, on one hand, and a more efficient fight against organised crime amongst the latter group, on the other hand, are prerequisites for more democratic and just societies, as well as a strengthened regional cooperation.

This proposed hypothesis will therefore provide the underlying premise for my comparative study of rearranging the puzzles of security in the Western Balkans. More specifically, the case studies of this hypothesis will encapsulate Kosovo, Serbia and FYROM.

The overarching aim of this research will be to:

Obtain insight and analyse the role of intra-state armed groups in Kosovo, Serbia and FYROM in intra- and inter-state democratisation and security.

Within this aim there are several objectives, namely to:

Analyse the reasons for the emergence of intra-state armed groups. What was the 1990s political situation preceding the conflicts in Kosovo, FYROM and Serbia?

What was the level of repression of ethnic Albanians in the region, and what was the rapport of the Serb and Macedonian communities with the former?

Examine the inter-connections and relations between intra-state armed groups across the borders of FYROM, Kosovo and Serbia. What was the contribution of the KLA in the formation of the other two armed groups? What were the goals, real or propagated, of these groups?

Identify their common denominators, their role in peace accords, and the challenges of the former combatants' integration in times of peace. Was the DDR and Security Sector Reform or Development (SSR/D) inclusive, fair, transparent?

Study the role of such groups in contributing to the implementation of peace agreements in partnership with their respective countries and/or international interlocutors. To what degree are these agreements observed vis-à-vis the international community partners (Kosovo case) or vis-à-vis Serbian and Macedonian governments?

Look into the external factors of influence. What was the role of international groups or agencies in supporting undercover activities of armed groups?

Look critically into the criminal elements within these intra-state armed groups and deduce their impact for the regional networks of organised crime. Is organised crime a regional challenge, since the cooperation in this area is inter-state?

The study is underpinned by a theory-driven research on security and social policy, taking as its basis an interdisciplinary approach (Zedner 2003).

The choice of location in South East Europe, more specifically the Western Balkans, is a geo-strategic one. The former Yugoslav states have undergone a specific experience of conducting wars in the 1990s, to building democratic societies and simultaneously consolidating security mechanisms. The ensuing security paradoxes, I believe, offer research findings that will inform similar experiences in other parts of the world. Moreover, the research has implications for security policies in western democracies, and produces valuable lessons learned in assisting other and newer post-conflict/transitional societies, or emerging democracies.

5 Conclusions

The conclusions of the research pertain to the question of integration of former combatants of intra-state armed groups in post-conflict transitory societies. What results from the dynamics of state *versus* former combatants turned politicians is the need for a more balanced, inclusive and constructive approach of the latter into the implementation of peace accords. Additionally, tackling the issue of deviant and/or criminal elements in the midst of armed groups is necessary for strengthened democratic societies. The end result of a fairer power in societies in Kosovo, FYROM and Serbia leads to a transformed security sector, it intensifies democratization processes, and it eventually helps the region get closer to European integration.

References

Allcock, John B. (2000) Explaining Yugoslavia. London: Hurst & Company

Banac, Ivo (1984) *The national question in Yugoslavia: origins, history, politics.* Ithaca, N.Y.: Cornell University Press

Bekaj, Armend R. (2010) *The KLA and the Kosovo war: from intra-state conflict to independent country*. Transition Series No. 8, Berlin: Berghof Conflict Research

Berdal, Mats and David M. Malone (eds) (2000) *Greed and Grievance: Economic Agendas in Civil Wars*. London: Lynne Rienner Publishers

Bianchini, Stefano & George Schöpflin (eds.) (1998). *State building in the Balkans: dilemmas on the eve of the 21st century*. Ravenna: Longo Editore

Bilandžić, Dušan. (1978) Historija Socijalističke Federativne Republike Jugoslavije – Glavni procesi. Zagreb: Školska knjiga

Biserko, Sonja (ed.) (2007). Human Rights – Hostage to the state's regression: Annual report, Serbia 2006. Belgrade: Helsinki Committee for Human Rights in Serbia

Clark, Howard (2000) Civil Resistance in Kosovo. London: Pluto Press

Clewlow, Ade (2010) *The Kosovo Protection Corps: A Critical Study of its Deactivation as a Transition.* Norwegian Institute of International Affairs. Security in Practice 4

Čolović, Ivan (2002) The politics of symbol in Serbia: Essays on political anthropology. London: Hurst & Co

Danforth, Loring M. (1995) *The Macedonian conflict: Ethnic nationalism in a transnational world*. Princeton: Princeton University Press

Di Lellio, Anna (2006) *The Case for Kosova – The Passage to Independence*. London: Anthem Press

Djokić, Dejan & James Ker-Lindsay (eds) (2011) New Perspectives on Yugoslavia. Key Issues and Controversies. London: Routledge

Đodan, Šime (1990) Yugo-tragedija, Zagreb: ST Komunikacije

Elshani, Gafurr (ed) (1998) Ushtria Çlirimtare e Kosovës. Dokumente dhe Artikuj. Botimi i dytë. Aarau: Zëri i Kosovës

Ellis, Burcu Akan (2003) *Shadow genealogies: Memory and identity among urban Muslims in Macedonia*. New York: East European Monographs

Ignatieff, Michael (1993) Blood and Belonging: Journeys Into the New Nationalism. New York: Farrar, Straus and Giroux

Judah, Tim (2000) Kosovo: War and Revenge. Yale: Yale University Press

Kaldor, Mary (2007) Human Security – Reflections on Globalization and Intervention. Cambridge: Polity Press

Kraja, Mehmet (2003) Mirupafshim në një luftë tjetër. Prishtinë: Rozafa

Lederach, John Paul (2002) *Building Peace: Sustainable Reconciliation in Divided Societies.* Washington D. C.: United States Institute of Peace Press

Little, Alan & Laura Silber (1995) Yugoslavia: Death of a Nation. London: Penguin

Malcolm, Noel (1998) Kosovo: A Short History. New York: New York University Press

Miall, Hugh, Oliver Ramsbotham and Tom Woodhouse (2003) *Contemporary Conflict Resolution*. Cambridge: Polity

Nordstrom, Carolyn (1997) A Different Kind of War Story. Philadelphia: University of Pennsylvania Press

O'Neill, William G. (2002) Kosovo: an unfinished peace, London: Lynne Rienner Publishers

Paris, Roland (2001) *Human Security – Paradigm Shift or Hot Air?* International Security, Vol. 26, No. 2, Fall, pp. 87 – 102

Ramet, Sabrina P. (2002) Balkan Babel – The Disintegration of Yugoslavia from the Death of Tito to the Fall of Milošević. Oxford: Westview Press

Ringdal, Kristen, Simkus, A., Listhaug, O. (2007) *Disaggregating public opinion* on the ethnic conflict in Macedonia. International Journal of Sociology 37: 75-95

Tanner, Samuel (2007) Political Opportunities and Local Contingencies in Mass Crime Participation: Personal Experience by Former Serbian Militiamen, Global Crime, Vol. 8, No. 2, May, pp. 152-171

Thachuk, Kimberley (2005) *Corruption and International Security*, SAIS Review, Volume 25, No. 1, Winter-Spring, pp. 143-152

Weller, Marc (1999) The Crisis in Kosovo 1989-1999: From the Dissolution of Yugoslavia to Rambouillet and the Outbreak of Hostilities. Cambridge: Documents and Analysis Publishing Ltd.

Zedner, Lucia (2003) The Concept of Security: an Agenda for Comparative Analysis, Legal Studies, pp. 153-176

Between national solidarity and local interests – The pro-governmental political orientation of the Romanians in Hungary (the end of the 19th century – the beginning of the 20th century)

Iudean Ovidiu Emil¹,

¹ PhD candidate, The Modern History Department, The Faculty of History and Philosophy, "Babeş - Bolyai"University of Cluj-Napoca, Romania, ovidiu.iudean@gmail.com

Abstract. The signing of the Ausgleich in 1867 had profound effects on the political situation of the nationalities inhabiting the Transleithan part of the Double Monarchy. For half a century, the Hungarian governments in Budapest approached these nationalities by means of a gradual magyarization policy, with different degrees of intensity.

The nationalities in Hungary reacted to this policy in three different ways. In the case of the Romanians, their political beliefs found expression within three major orientations: national-passivism, national-activism and governmental-activism.

This paper aims to offer an analysis of the latter political orientation (governmental-activism) by outlining its main characteristic elements, and comparing it to the other two Romanian political orientations. Furthermore, the key figures of this orientation will be identified and an analysis of their activity will be supplied.

From a methodological perspective, two major approaches will be employed. Firstly, the pro-governmental political direction will be theoretically analysed and the findings will be corroborated with the factual presentation of its key events. To complement this approach, this paper shall also provide a comparative analysis between governmental-activism and the other political directions in Hungary.

This political direction has been less discussed in the Hungarian historiography and relegated to the periphery by Romanian historians as a result of the "renegade" image associated with its adherents in contemporary scientific environments and during the half of the 20^{th} century. Therefore, the main strength of this

paper lies in the fact that it will provide a novel approach towards this marginalized but nevertheless significant orientation, attempting to fill the historiographical gaps in the subject.

Keywords: Austro-Hungary, the Romanian national movement, political currents, Romanian governmental representatives, electoral strategies

1 Introduction. General political background

The establishment of the Austro-Hungarian monarchy as a result of the political compromise between the Austrian and the Hungarian leaders – the only solution to safeguard the monarchy in the context of its political crisis - meant significant changes for the nationalities inhabiting its territories. The situation of the nationalities within Budapest's area of government drastically changed: an involution from the status of local autonomies granted by the legislation passed during the Austrian empire's "constitutional" period (1860-1867) occurred. Owing to the Ausgleich, these territories were directly subordinated to the Hungarian government. The relations between the nationalities and the central Hungarian authorities were greatly affected by the latter's magyarization policy, a constant of all Budapest governments which oversaw the Transleithan part of the Double Monarchy for half a century. Despite its constancy, the implementation of this policy was gradual, and the intensity of each stage greatly varied. For instance, at the beginning of the dualist period a series of laws which limited the rights of the nationalities compared to the earlier periods but provided them with some concessions (The Nationalities' Law of 1868) were passed. However, during the latter half of the 19th century and the beginning of the 20th century a set of laws was passed (The Trefort Law of 1883, The Apponvi Law of 1907, etc.) which deeply affected the nationalities' cultural, educational and political development.

The beginning of Kálmán Tisza's long government (1875 - 1890), which had the main goals of stabilising and solidifying the Austro-Hungarian construct, also marked the start of a new period from the perspective of the relations between the central Hungarian authorities and the nationalities. In the first decade since the Ausgleich was signed, the Hungarian political life was dominated by a series of personalities such as Ferenc Deák and József Eötvös, who maintained to a policy of cooperation with the nationalities in Hungary.

The nationalities in Hungary sought to react to the newly established political situation, but found themselves hindered by the lack of an experienced political class, and of a necessarily high social, cultural and economic level.

2 Literature Review

The political life of the Romanians in Hungary at the end of the 19^{th} and beginning of the 20^{th} centuries has been a constant preoccupation of the Romanian historiography. Many works and studies regarding this subject have been written in the period between the two wars or during the communist period. Even after 1989, when the area of available research topics was expanded as a result of the lack of censorship, the abovementioned subject remained an interesting one for Romanian historians.

Despite this fact, two issues have enjoyed a more constant attention: firstly, the key moments of the political life of the Romanians in Hungary and secondly, the analysis and interpretation of the two political currents which characterized the national (national-activism movement and national-passivism). The activistgovernmental/moderate current has been rarely discussed by Romanian historians. With the exception of a low number of biographies of key personalities which adhered to this current, and its mentioning in general works which discuss Romanian political life in the abovementioned time period, the study of this third Romanian political direction has been relegated to the periphery of the Romanian historiography. One explanation for this phenomenon may be the negative image projected on the adherents of governmental activism during their period of activity as well as by posterity (during the interwar and communist periods). The political compromise between these adherents and the Hungarian governments – in the context of tense relations between nationalities and the central authorities – not only led to their ostracizing, but also determined the indifferent reaction of the Romanian historians towards this subject, both in the past and presently.

Similarly, the Hungarian historiography has focused on the general research of the political life in the territories inhabited by the Romanians during the dualist period, emphasizing the way in which the Hungarian governments related to the nationalities. In this matter, Hungarian historians have underlined the substantial role played by the nationalities in ensuring the electoral success of the ruling Hungarian party, through their contribution to the obtaining of parliamentary majority.

It is precisely because this paper deals with a subject that has not been properly discussed in historiography that a wide array of historical sources will be employed, with an emphasis on archival documentation and the Romanian 19th and early 20th century press which was contemporary to the events presented and analysed. The material from the four archival funds that have been explored, corroborated with the information obtained from the Romanian newspaper articles, will be complemented with general reference works from the Romanian and Hungarian historiographies.

3 Data and Methodology

Taking into account the specific issues arising from the discussion of a less researched subject, this paper will attempt to provide a theoretical analysis of the three political orientations within the national movement of the Romanians in Hungary, with a special emphasis on the pro-governmental current.

Following this theoretical approach, a quantitative and qualitative analysis of the Romanian activist-governmental faction in Hungary will be provided, in order to present its main political representatives, the degree of their involvement and their role in the Hungarian political life. A comparatively large section of this study will be dedicated to the analysis of the electoral involvement of the adherents to governmental activism. This aspect will be treated by focusing both on electoral strategies and governmental candidatures, as well as on the behaviour of their Romanian electoral supporters.

4 Empirical Analysis

4.1 The political life of the Romanians in Hungary

In the case of the Romanians in Hungary, who amounted to approximately 15 % of its population, several different political orientations were designed and employed by their political leaders, who attempted to adapt them to the geopolitical realities of their respective territories. Thus, in the first years following the signing of the *Ausgleich*, the Romanians organized themselves into two political parties according to the historical region they inhabited. Consequently two major parties were established: firstly, the National Romanian Party of Transylvania (NRPT) which coordinated the political life of the Romanians living within the Carpathian arch; secondly, the National Romanian Party of Banat and Hungary (NRPBH), which coordinated the political life of the Romanians outside the Carpathian arch, who inhabited an area to the West and South-West of the abovementioned region.

The political orientations adopted by these two political organizations differed from their inception, the major factor behind this differentiation being the provisions of the electoral Hungarian law. Thus, for the intra-Carpathian space the provisions of the abovementioned law created a climate which undermined the Romanian nation's possibilities to become involved in Hungary's political life in a manner proportional to their numbers. The extremely high census and the geography of the electoral districts limited the role of the Romanians in the electoral process. Consequently, the NRPT adopted the tactic of national-passivism, abstaining from the appointing of its own candidates and from the participation in the electoral process, as a sign of protest against the political-electoral realities of this area.

As far as the extra-Carpathian Romanian-inhabited territories (Banat, Crişana, and Maramureş) are concerned, they enjoyed the same electoral provisions which applied to the entire Hungarian territory (with the exception of Transylvania). As a result of these provisions, and especially that referring to the lower census, the Romanians in these regions could express their political opinions more adequately. The natural consequence was the adherence of the NRPBH to the political tactic of national-activism.

After a decade and a half of political activity in the Transleithan part of Austro-Hungary, which was characterized by an extremely low amount of accomplishments in the field of national politics, the leaders of the Romanians realized the necessity of creating a single organism that would coordinate their political life in a unitary and coherent way. Thus, the decision to unite the two Romanian parties was reached at the National Conference in Sibiu in 1881, and the National Romanian Party (NRP) was created. Its political orientation was a hybrid between activism – in the extra-Carpathian areas – and passivism – in the intra-Carpathian areas. Only in 1887 was national-passivism adopted in the entire Romanian-inhabited territory and national-activism renounced, as a result of the intensification of the Hungarian government's policy to electorally subordinate the Romanian counties.

The political programme of the newly-established national party was created during the national Conference that took place in Sibiu in May 1881. This was divided into nine programmatic points. The first referred to the party's intention to militate in favour of the regaining of Transylvania's autonomy (the intra-Carpathian territory inhabited by Romanians), which had been lost at the establishment of the Austro-Hungarian dualism. Owing to the inclusion of this provision, the NRP's programme overtly contradicted the terms of the *Ausgleich* and the notion of a unitary state. The other provisions concerned the following issues: the use of the Romanian language in administration and justice within the Romanian-inhabited territories; the increase in the number of Romanian clerks employed by the state; the modification of the 1868 Law of Nationalities and its appropriate application; the protection of the Romanians' ecclesiastical and educational rights and the extension of the right to vote.

Significant modifications were brought to it only in 1887, when passivism was extended to the entire territory inhabited by the Romanians, and in 1905, when the Romanians' political orientation was drastically changed through the implementation of national-activism. For more than 35 years, the NRP adopted an attitude that directly contradicted the Budapest governments' magyarization policy. The NPR leaders believed that, due to the political requirements they had formulated in the national political programme, they could not adopt a position similar to that of the Hungarian governments.

The political divergences between the NRP and the government culminated in the moment of the so-called "memorandist" action at the beginning of the 1890s. Having adhered to passivism, the Romanians' national leaders relied on the writing of petitions which listed their claims (*petitionalism*) as a political tactic. Thus, it was decided that a document entitled "Memorandum" would be created and personally delivered to the emperor-king Franz Joseph I.

Franz Joseph I considered that the matter of the Romanians' requests did not fall under his jurisdiction and consequently sent the document, unopened, to the Hungarian government in Budapest. This authority returned the sealed document to the NRP. The Hungarian authorities believed that by requesting national autonomy, the authors of the "Memorandum" had actively contested the validity of the union between Transylvania and Hungary. In addition, the dissemination of such a document was considered a crime, for which the authors were tried. At the conclusion of the trial in May 1894, 14 leaders of the NRP were sentenced to incarceration. In January 1895, the NRP was dealt a new blow by the Hungarian authorities, when it was outlawed through a decree of the minister for internal affairs. The negative effects of the *memorandist* movement led to an internal crisis in the NRP, which persisted until 1905, when the NRP began a new stage in its political evolution, that of nationalactivism. The divergences between the Hungarian authorities and the NRP allowed for the assertion of a category of Romanian politicians who embraced moderate views and were willing to collaborate with the Budapest government. In the beginning of the 1880s, the idea of establishing a new political party, with a distinct program, which would promote the collaboration between the Romanian and the Hungarian nations in order to better the former's social, cultural and economic conditions, came to the fore. This idea originated within a significant segment of the Romanian political leadership, in the context of the NRP's discussions regarding the appropriate tactic for the newly-established party. This type of collaboration (which was rejected by the NRP) was seen as viable only if the Romanian political leaders who envisaged cooperation managed to obtain the support of the Liberal Party (*Szabadelvű Párt*), which was governing Hungary. This type of collaboration with the central authorities was not a novelty from the perspective of the Romanians' political life in Hungary: several Romanian leaders had adopted this type of attitude in the first decade and a half since the signing of the *Ausgleich*, without being politically organized.

After a first failed attempt in 1881, the Romanians who had activist-governmental views managed to organize themselves into a new party before the parliamentary elections of 1884. The party's meetings were only permitted during the electoral campaign period, as per the provisions of the XXXII/1874 Electoral Law. The MRP was established as a result of the conference of the Romanians in Budapest (March 1884), which also designed and adopted its political programme, elected the members of its central committee, and was presided over by the Romanian Metropolitan of Sibiu – the highest hierarchical level within the Romanian Orthodox Church in Hungary. The Romanian representatives in the Budapest Parliament and administrative county clerks were found among other members of the central party committee. The party was self-dissolved in 1885. Negotiations were held between the Hungarian political parties and the leaders of the moderate-governmental faction in order to re-establish this party at the beginning of the 1890s - before the 1896 elections – and at the end of the first decade of the 20th century – before the 1910 elections. Even after 1905, when the NPR reinstated its activist orientation, the adherents to governmental activism still won representative offices on behalf of the Romanians in the Budapest Parliament.

The continuity of the idea of a moderate party can be explained thusly: the late 19th century and early 20th century Hungarian governments felt a strong necessity to collaborate at a political-electoral level with the moderate elements from within the other nationalities in order to obtain the highest possible number of representative offices in those counties which had a majority of non-Hungarian inhabitants. Such compromises were established before parliamentary elections between the central authorities in Budapest and certain political factions which held pro-governmental views, as for instance certain leaders of the Serbians or the Saxons in Hungary. The efficiency of such an approach is evidenced in the high number to offices obtained in the areas inhabited by other nationalities (approximately 2/3 of the Hungarian ruling party's representatives were elected in electoral districts where the majority of the population was not of Hungarian nationality).

4.2 Definition of the activist-governmental current and its main characteristics

Governmental-activism was one of the three political currents to which the Romanians in Hungary adhered during the dualist period (1867-1918). This current was characterised by a series of elements that manifestly distinguished it from the other two Romanian orientations - national-activism and national-passivism. The main distinguishing characteristic was its acceptance of the political realities extant within the Double Monarchy as a result of the signing of the Ausgleich, particularly the union between Transylvania and Hungary and the resulting loss of the province's autonomy. Governmental-activists believed that only by placing themselves in a constitutional framework would they be able to organize their political activity without constantly being the object of the state's judicial inquests and denunciations by the Hungarian political class. In this respect, the activist-governmental movement was characterized by openness to collaboration with the Hungarian government, an element which was seen as fundamental for the accomplishment of their political goals. It was believed that only through a policy of compromise would the Romanian nation be able to evolve educationally, culturally and economically, and to protect the ecclesiastical rights it had obtained during the liberal period (1860-1867).

More exactly, the movement's representatives vowed to adopt a supportive attitude in regards to the Hungarian governments' state-wide political projects, in exchange for certain concessions at a local, county level. One of the main distinguishing elements of governmental-activism was precisely that of the political approach towards the national Romanian "question" in Hungary: while the NRP favoured the fulfilment of nation-wide objectives, the priority of governmental-activists was the resolving of localised issues. It should however be mentioned that governmental-activists as well as the adherents of national-activism and national-passivism were similarly preoccupied with the development of the entire Romanian nation in Hungary. The difference between these two currents was visible however in regards to the direction that the proposed betterment of the situation of the Romanians should take: from general (the Romanian nation) to local (the Romanian communities within each county) – the NRP's objective; from local to general – the governmental-activists' objective. In the latter's opinion, beneficial local developments at a community level would have eventually been reflected in the state of the entire nation. Moreover, these objectives were seen as easier to achieve, due to the Hungarian authorities' greater availability to make concessions for smaller areas. This greater availability stemmed from the fact that such policies did not contradict the Hungarian governments' general objectives and also protected them from vehement reactions from the opposition parties, which militated for the acceleration and amplification of the magyarization process.

To a certain extent, the governmental-activist movement had also gained the approval of the Bucharest governments, which were interested in the improvement of the political relations between the Kingdom of Romania and the Austro-Hungarian Monarchy. A release of tension in the relations between the Budapest governments and the Romanian nation in Hungary was seen as beneficial to the accomplishment of the Romanian Kingdom's external policy objectives.

4.3 Political orientations within the movement

The governmental-activist movement had a gradual evolution, and was heterogeneous. Within it three different political orientations coexisted, during the entire dualist period: that of the governmental-nationals, that of the governmental-moderates and that of the governmental-'renegades'.

The least represented group was that of the *governmental-nationals*. This was constituted of those elements that accepted the necessity of collaboration with governmental authorities in Hungary in order to accomplish national Romanian objectives, but were not open to compromising on the issues of the national programme, and vehemently reacted to any action of magyarization. This orientation was closest to the national-activism promoted by the NRP, as its name indicates. Only a limited number of representatives of the activist-governmental movement adopted this type of political orientation. The Hungarian government gradually refused to collaborate with the representatives of governmental-nationalism and as a result, its adherents left the activist-governmental movement.

The second orientation or group - that of the *governmental-'moderates'* - was more in tune with the general characteristics of the movement. Its adherents had constantly manifested their support for the governments' political programmes, while in exchange for this support they had requested both local-level and nation-wide concessions for the Romanians in Hungary. They attempted to temper the process of magyarization by suggesting compromise solutions or by requesting concessions in other domains which were essential for the well-being of the Romanian nation. The governmental-moderates laid the foundations of the only party that reunited the adherents of the activist-governmental current in a singular political organisation, distinct from that of the governing Hungarian party - in the ranks of which they were always included. The structure of the new party was exclusively Romanian. In some periods of the dualist epoch, the governmental-moderates manifested their openness to a potential collaboration with the NRP national-activist branch, in order to accomplish common goals.

The so-called *governmental-'renegades'* comprised the third orientation within the movement, which distinguished itself through a complete openness to the collaboration with the Budapest governments. They had local-level objectives and were strictly preoccupied with the improvement of the Romanian nation within those administrative units which they represented or inhabited. Their goal was to obtain educational, cultural and social concessions without negatively reacting to the authorities' magyarization tendencies. An eloquent example of their attitude in this respect is that regarding the issue of the Romanians' educational improvement. They militated both for an increase in the number of state-institutions providing education in the territories inhabited by Romanians, and the improvement of the level of education offered by these institutions. Their request was that of free and unrestricted access to these institutions for young Romanians. They believed that, despite the fact that courses of study were offered only in Hungarian, the possibility of receiving an education (ranging from primary to secondary and university level) represented a fundamental gain for the Romanians in Hungary.

4.4 The stages of development of the activist-governmental current

Within the development of the activist-governmental current three stages can be identified, as a consequence of the pre-eminence of one or another of the abovementioned orientations: the stage of the forerunners, that of the 'moderates' and that of the 'renegades'.

The forerunner stage (1867 – 1881) can be placed chronologically between the signing of the dualist agreement and the year when the creation of a Moderate Romanian Party was first attempted. During this stage the three orientations found a quasi-balanced expression in the context of a relatively confused Romanian political class. The main issue was that of finding a suitable approach for the new political realities created by the establishment of the Double Monarchy. Varied solutions were suggested, due to the fact that the Romanians' political organisation was still in an incipient phase, characterised by the disunity between the actions of the intra-Carpathian and the extra-Carpathian political class. Moreover, during this period the Romanian political environments were dominated by the intensely-promoted idea that the *Ausgleich* was ephemeral, and that a return to the previous liberal state was inevitable. In this context, it was difficult to clearly and precisely frame a Romanian politican's activity within a specific current. This attempt was further hindered by the increased and intense mobility between political circles.

There were however Romanian political leaders who promoted solutions which were close to those proposed by governmental-activists. Two of the most remarkable representatives of this category of leaders were the Orthodox Metropolitan Andrei Şaguna and the magnate Emanuil Gojdu. Both militated for collaboration with the Hungarian authorities for the benefit of the Romanian nation. Their Hungarian homologues were equally distinguished politicians, such as József Eötvös and Ferenc Deák, who shared their ideas and believed that Romanian-Hungarian cooperation would be advantageous to both nations. It should also be noted that during the entire dualist period, the Hungarian governments that supported the Romanian activist-governmental current emphasized the need to promote this cooperation in the context of the geopolitical realities that had placed the Hungarians and the Romanians in the midst of several Slavic nations.

The stage of the 'moderates' (1881-1905) can be placed, from a chronological perspective, between the moment when the first projects to establish a Romanian activist-governmental party were devised, and the moment when the NRP adopted national-activism. During this period, the 'moderates' dominated not only the activist-governmental current but also the entire political life of the Romanians in Hungary from the perspective of parliamentary activity. They were superior from a quantitative standpoint and managed to successfully impose their political ideas. In the above-mentioned timeframe, during 7 parliamentary cycles, 65 representative mandates were won, averaging over 9 mandates per electoral cycle. If the fact that the Budapest Parliament was comprised of 413 elected representatives is taken into consideration, the percent of activist-governmental mandates only slightly exceeds 2%. Despite this limited representation, the number of representatives was sufficient to support the political tactic of the Romanians' situation at a local level that would eventually improve the nation-wide situation.

For a clearer image regarding the pre-eminence of the 'moderates' over the other

867

two orientations between 1881 and 1905, it should be mentioned that, from the total of 65 mandates obtained by governmental-activists, more than 80% were held by this orientation's adherents. Moreover, the 'moderates' were also supported by the Orthodox Metropolitans of Sibiu, Miron Roman and later Ioan Metianu, and by an important segment of the Romanian Orthodox clerics. In the initial years of this stage, the governmental-activists attempted to and were partially successful in collaborating with the activist-nationals. A result of this common action was the establishment of the Romanian Moderate Party (1884). This organisation was dominated by the adherents of the moderate orientation within the governmental-activist orientation, but also comprised adherents of national-activism and even some elements of the governmental-'renegades'. Their political platform was not significantly different from the NRP's programme. The first point of the former platform specified that the RMP's members were under the obligation to conform to the provisions of the Hungarian constitution. Thus, the idea regarding Transylvania's autonomy was abandoned and the 1867 Ausgleich was recognized. A significant distinction could be observed in the matter of electoral provisions. The NRP's programme promoted the introduction of the universal vote, while the RMP's platform only supported an extension of the Hungarian electoral provisions to the intra-Carpathian territory inhabited by Romanians (Transylvania). The other programmatic provisions contained goals similar to those expressed by the NRP.

However, the RMP's existence was brief. Shortly after the conclusion of the 1884 parliamentary elections, the Hungarian government retracted its support for the party leaders. The goal of the government presided by Kálmán Tisza, that of obtaining of a parliamentary majority in order to implement his political projects, had been accomplished. This was made possible by the nationalities' wide support, among which the Romanian voters with governmental-activist views were extremely significant. The majority of the Magyar Liberal Party's mandates were won in those electoral circles in Hungary that were preponderantly inhabited by non-Hungarians. In the circles where the Hungarian population was in majority, the opposition parties managed to win most of the parliamentary seats. This type of situation was a constant of the Hungarian political life during the entire span of the dualist regime. The governmental party won the majority of its parliamentary seats in the electoral circles dominated by non-Hungarian voters. This explains the Hungarian political leaders' receptiveness towards and support of the Romanian activist-governmental current. In exchange for the Romanian voters' pro-governmental attitude, the government offered the Romanian activist-governmental candidates between 1 and 3 representative mandates per county. These were offered mostly in those counties from the extra-Carpathian regions inhabited by Romanians, where the activist-governmental current was especially well received.

Despite the party's dissolving at the beginning of 1885, several projects to reestablish it were devised in the mid-1890s (during the NRP's crisis provoked by the memorandist action) and in the early 20th century (in the context of the Hungarian parliamentary crisis). These attempts took place preponderantly during the parliamentary elections, were based on the 1884 political platform created in Budapest and gathered relatively the same Romanian political circles. *The 'renegade' stage* (1905-1918) was the final period of the Romanian activistgovernmental movement. It was chronologically delimitated by the moment when the NRP's leaders adopted national-activism and the end of the Austro-Hungarian monarchy as a result of the First World War. During this stage, the adherents of the activist-governmental current were characterised by a greater openness towards the Hungarian governments' policies, and by mainly localised goals. This stage was dominated by the representatives of the so-called governmental-'renegade' group.

These were named 'renegades' because they were placed in direct opposition to the NRP's activist-national representatives by the contemporary nineteenth–century Romanian public opinion. With the exception of national-activism any other political current was seen by an important part of the Romanian public opinion as anti-Romanian. The NRP was supported by the majority of the Romanian newspapers in Hungary, and consequently, a wide-ranging and significant press campaign directed against the adherents of the activist-governmental current, and particularly against the 'renegades' was organized. Taking into account the major influence of the articles which appeared in the national press, the reaction of the Romanian public opinion towards the activist-governmental current was understandable.

During the 1905, 1906 and 1910 parliamentary elections, the national-activist candidates managed to win over 25 mandates, while at the 1906 and 1910 elections, the activist-governmental candidates only won 14 mandates, 10 of which were obtained by the adherents of the 'renegade' orientation. On numerous occasions, the adherents of both currents competed against each other within the same electoral unit. This determined the aggressive reaction of the Romanian press against the activist-governmental candidates during the electoral campaigns. However, these candidates were continually supported at a local level, a fact which is especially reflected, for instance, in the intermediary elections that took place in 1909 in the electoral circle of Oravita. During these elections, the NRP president, George Pop of Băseşti ran against one of the main representatives of the governmental-'renegade' orientation, Iosif Siegescu. The mandate was won by the activist-governmental candidate with a striking majority: 66, 46% of the votes went to Siegescu, while only 33, 54% of the votes went to Pop of Băseşti.

4.5 The political representatives of the activist-governmental current

The political representatives of the current can be divided into two major categories: the *secular elite*, which comprised the parliamentary deputies, and the *ecclesiastical elite*, which comprised the Romanian Orthodox and Greek-Catholic metropolitans and bishops.

The activity of the latter category was less visible from the perspective of the direct involvement in the activist-governmental actions. However, they expressed their viewpoints during the sessions of the Superior Chamber of the Budapest Parliament, and also distinguished themselves through their attitudes during the electoral campaigns. This segment of the elite constantly militated for the extension and protection of the ecclesiastical and educational rights of the Romanians in Hungary. During the elections, they expressed their support for the government candidates, especially for those of Romanian origin. As a concretisation of their beliefs they sent out encyclicals towards priests, encouraging them to support the governmental candidates and to capitalize on their influence over the Romanian voters, especially in the rural areas.

Their position in the ecclesiastical hierarchy and their financial dependence on the state authorities made them adopt this type of position, in accordance with the ideas of the activist-governmental current.

The most visible representatives of the current were the Romanian activistgovernmental deputies in the Budapest parliament. Between 1881 and 1918 they won no less than 79 mandates, although they never held more than 2,5% of the total number of seats during a parliamentary cycle. They based their political activity on the characteristic ideas which have been outlined in the previous subchapters. However, they did not distinguish themselves through a vast parliamentary activity. During the 9 parliamentary cycles in the above-mentioned period, the activist-governmental representatives rarely held speeches or intervened in the debates. They were however part of several parliamentary committees, and in this quality were often able to resolve individual issues or community matters for the Romanians at a local level. It is in this aspect that the results of their political activity are most visible. They distinguished themselves through the following: the directing of governmental funds intended for the building of secondary-level educational institutions towards the Romanianinhabited areas; the extension of the railway lines in the same territories; the obtaining of licences for the practice of different trades and the release from the payment of taxes for their co-nationals; the awarding of scholarships for young Romanians; the protection of the interests of the Romanian Greek-Oriental and Greek-Catholic Churches.

The Romanian activist-governmental representatives were recruited from the main Romanian socio-professional categories: lawyers, professors, clerics, administrative clerks. What is especially interesting in this respect is the fact that all three Romanian professors who chaired the department of Romanian language and literature studies at the University of Budapest from its establishment in 1863 until 1918 were activistgovernmental representatives.

4.6 The adherents of the activist-governmental current

In regards to the adherents of the activist-governmental current, a multilevel analysis must be made, focusing on the following issues: their social background, the reasoning and justification for the adoption of this current's principles, and their political and electoral behaviour.

Firstly, it must be noted that this current spread among all of the Romanian social categories in Hungary. During the second half of the 19th century most of the Romanians in Hungary were *small landowners*, whose base occupation was agriculture. Only a limited part of the Romanian nation was middle-class. These constituted the so-called *Romanian intelligentsia*, namely those persons who had benefitted from a secondary- and university-level education and who were employed as priests, professors, teachers, lawyers and administrative clerks. The Romanian intelligentsia was especially influent in the rural environment, where the population's overall level of education was extremely low.

Despite the fact that some differences in regards to the motivation behind their adherence to the principles of the activist-governmental current can be observed when comparing the two categories of voters, their behaviour is similarly framed. The elections were seen by a considerable part of the Romanian electorate as a means of obtaining different types of benefits. Most of these benefits were financial: for instance the bribes offered by governmental electoral agents in exchange for votes given to the candidate they represented – a wide-ranging practice in dualist Hungary. The amounts offered varied between 1 crown and 100 crowns, depending on the voter's importance within the local community and his degree of influence. Other benefits comprised personal favours for voters, as for instance the expediting of retirement procedures, the appointment in administrative offices, the concession of the rights to the exploitation of natural resources, the awarding of scholarships for voters' sons, etc. Potential benefits for local communities included: the granting of the right to pasture on a certain area, the release from the payment of taxes for a period of time, etc. It is therefore possible to discuss these motivations in terms of an electoral opportunism, that was characteristic to the adherents of the activist-governmental current and that was manifest during the entire period between 1881 and 1918.

The voters' electoral behaviour was also influenced by the pressures coming from the Hungarian authorities, who were preoccupied with the promotion of the governmental candidates' interests. The most vulnerable voters to this type of pressures were the socio-professional categories who were directly depended on the Hungarian state: the priests, the teachers and the state clerks. The refusal to adopt an appropriate electoral attitude in accordance with the government's wishes could lead to unwanted consequences for these categories. Most clerks risked losing their jobs, while priests and teachers could be deprived of state financing, which formed the basis of their annual incomes. Among the small landowners, pressures were likely to turn into the oftentimes violent abuses of the local authorities. Therefore, the second significant component of the activist-governmental electoral behaviour was characterized by corruption, abuses and the electoral fraud of the governmental candidates' agents.

5 Conclusions

In the final section of this paper, we will attempt to draw a series of conclusions concerning the reasons which have led us to see the political action of the activist-governmental representatives in terms of a distinct current which was well represented at the level of the Romanian nation in Hungary.

Starting from the opinion that governmental activism did not represent a nichecurrent in the general framework of the political options of the Romanians in Hungary, we will attempt to infirm the arguments advanced by the Romanian historiography of the interwar, communist and even contemporary periods, which claimed the opposite.

Firstly, one of the arguments that previous historians had advanced was the fact that there was a very limited number of parliamentary representatives that had won their mandates by running on a activist-governmental programme. Concerning this aspect, it should be noted that the average of number of mandates won by the activistgovernmental representatives for each parliamentary cycle is very close to that of the mandates won by the NRP's adherents after national-activism was adopted. Therefore, the parliamentary representation of the two currents was at least similar, if not equal. It must also be noted that during the entire period between 1887 and 1903, the activistgovernmental deputies constituted the Romanians' only delegates in the Budapest Parliament. On the other hand, the NRP's political activity during this time was more or less theoretical. Consequently, this activity lacked the expected results for the betterment of the situation of the Romanian nation in Hungary.

Secondly, it was also stated that there was a lack of support from adherents that would legitimize this current. This issue should be better nuanced. The Romanian voters constantly supported this current during the entire dualist period. We cannot however disregard the factors on which this support was based. It is an undeniable truth that, behind this endorsement were the voters' opportunistic motivations and the electoral pressure exerted by the Hungarian authorities, within the wider context of a Romanian nation which lacked an adequate electoral organisation. Despite this fact, Romanians who had the right to vote accepted the existing situation and learned to use it to their best advantage. They weighed the political options they were given by the Romanian political class and concluded that by adhering to the activist-governmental current they could contribute to the cultural, education and economic development of the Romanian nation.

The Romanian historians have refused to see the adherents of the governmental type of activism as representatives of a third political current. This opinion has been justified by the assertion that the adherents of governmental-activism denied their Romanian ethnic origins by adopting a position which seemingly contravened that of the national movement, represented by the NRP. In order to contest this opinion, we may bring as proofs the electoral speeches held by the activist-governmental candidates. They constantly declared their ethnic Romanian background as well as their intention to act for the benefit of the Romanian nation. A denial of their ethnic background would have shown a complete lack of political tact, as most of the voters in the electoral circles in which they had ran for office were Romanians. Moreover, despite the fact that they had won their mandates by running on the ruling Hungarian party's lists, the Hungarian politicians saw them as parliamentary representatives of the Romanian nation. These arguments are only two examples which can advanced in order to justify the need for a hystoriographical re-evaluation of the activistgovernmental current, in its Romanian manifestation.

The main goal of this study was to provide a broader perspective of the political events in Hungary at the end of the 19th century and the beginning of the 20th, from the viewpoint of the relations between the Hungarian governments and the Romanian nation, through the detailed discussion of a political current that has been marginalized by both the Romanian and the Hungarian historiographies, for different reasons. In addition, by studying the role that the political adherents to the activist-governmental current played in the political life of the Romanians in Hungary, we hope to have recovered an important component of a larger theme for the Romanian historiography, which will provide a more coherent narrative for the period and region in question.

References

Biblioteca Academiei Române, Arhiva personală Alexandru Roman [The Library of the Romanian Academy, Personal archive of Alexandru Roman]

Arhivele Naționale ale României Serviciul Județean Arad, *Fond personal Roman Ciorogariu* [The National Romanian Archives, Arad County Service, *Personal fund of Roman Ciorogariu*]

Viitorul, Budapesta, 1883-1885 [The Future, Budapest]

Telegraful român, Sibiu, 1881-1910 [The Romanian Telegraph]

Gazeta Transilvaniei, Brașov, 1881-1910 [The Transylvanian Gazette]

Brote, E (1895) Cestiunea română în Transilvania, Tipografia Voința Națională, București [*The Romanian question in Transylvania*, National Will Printing Press, Bucharest]

Gero, A (1997) The Hungarian Parliament (1867-1918). A Mirage of Power, Columbia University Press, New York

Hitchins, K (1999) A Nation Affirmed: The Romanian National Movement in Transylvania (1860-1914), Editura Enciclopedică, București [The Encyclopaedic Press, Bucharest]

Kontler, L (2009) A History of Hungary, Atlantisz Publishing House, Budapest

Ormos, M, Király, B (2001) *Hungary: Governments and politics*, Columbia University Press, New York

Păcățian, T (1913) *Cartea de Aur*, vol. VII, Tiparul Tipografiei Arhidiecezane, Sibiu [*The Golden Book*, vol. VII, Arch-diocesan Printing Press]

Popovici, V, Iudean, O (2011) The Elective Representation of the Romanians in the Hungarian Parliament. *Studia Universitatis Petru Maior. Historia*, 11/2011, 121-146

Seton-Watson, R (1911) Corruption and Reform in Hungary. A Study of Electoral Practice, Constable & Co. LTD, London

Teodor, P, Maior, L, Bocşan, N (1994) *Memorandul (1892-1894): ideologie şi acțiune politică românească*, Editura Progresul Românesc, București [The Memorandum (1892-1894): ideology and Romanian political activity, The Romanian Progress Press, Bucharest]

Toth, A (1973) Parteien und Reichstagwahlen in Ungarn 1848-1892, R. Oldenbourg Verlag, München

The Romanian Administrative Elite from Transylvania in the Liberal Period (1861-1867)

Alexandru Onojescu¹

¹ PhD. Candidate, Department of Modern History, Archive Studies and Ethnology, Babeş-Bolyai University, Cluj-Napoca, Romania alexandruonojescu@ubbcluj.ro

Abstract. The purpose of the essay is to analyze the structural evolution of the Romanian high civil servants corps from Transylvania during the liberal period of the Habsburg Monarchy (1861-1867). We intend to use a descriptive approach that will develop gradually into a 4-level investigation: 1^{st} . the social background of the Romanian appointed high civil servants of Transylvania (social origin, intellectual training) 2^{nd} . the recruitment mechanisms; 3^{rd} . the administrative career of these officials, and 4^{th} . their role in the shaping of a national political strategy for the Romanian national movement, in its attempts to establish itself among the other constitutionally recognized nations of Transylvania.

Keywords: Transylvania, elite, bureaucracy, Romanians, national movement

1 Introduction

The years between 1861 and 1867 brought some of the most interesting events for the political history of the Romanians under the Habsburg rule. The Great Principality of Transylvania, as well as other parts of the Habsburg Empire, reached a new political and administrative stage: the so-called liberal period, which marked the reinstatement of its full historical autonomy. From the perspective of the Romanian national movement, this period was characterized by an increasingly dynamic political and national process of demand. The Romanians requested for themselves a juridical-constitutional status equal to that of Transylvania's other peoples, on the basis of an argumentation drawn from history, demography and finance. A fundamental part of this struggle was the necessity for the Romanians to install as many of their representatives as possible in the administrative, political and juridical structures of Transylvania, both at a central and at a local level, as far as these institutions had begun to coalesce around 1861. Consequently, the favourable context enabled the number of Romanian civil servants to double by 1863, in comparison to the previous period (1848-1859), a fact that was easily noticed by contemporaries (Bariţiu 1891, 156).

Given the historical premises, those who entered the Romanian bureaucratic elite became not only clerks, but also representatives of their nation and denominations. Thus, they became entangled with the whole national political movement, a situation which turned back on them after the 1867 *Ausgleich*. Despite all these, the short time period in which they functioned atop the provincial administration brought many positive consequences, both for the Romanian nation and for them personally.

2 Literature Review

The first Romanian historiographical texts on the 1861-1865 period date back to 1890-1910 (Barițiu 1889-1891, Păcățian 1904-1905) and they represent an attempt to justify the Romanians' struggle for political and national emancipation. These writings, based on memoires and on the times' press, were written by semiprofessionals, were ideologically engaged and charachterized by an undeveloped critical spirit. The information they encompass are helpful, but many are incomplete or bring forward a unilateral perspective.

This trend carried on during the interwar period, the difference being that now, historians from the newly-formed Greater Romania attempted to justify the 19th century struggle for emancipation through the finality of the 1918 union.

In the Communist period, especially during the period of Communist Nationalism (1975-1989), studying the national movement became thorough, the liberal era, through its achievements, being regarded as an important moment of the political emancipation of the Romanians. The eulogistic presentation of the years 1861-1865 served as a comparison for the deprecation of the 1867 Ausgleich, which was regarded as the inception of the final stage of the Romanian nation's oppression. This dichotomic perspective was framed in a Marxist explanatory grid that emphasized the role of the Romanian bourgeoisie, backed up by the rural proletariat, on the historical stage (Netea 1974, Pascu 1978).

Unencumbered by the communist ideological impositions and limitations, the Romanian historiography concerning Transylvania began, after the 1990s, to focus on a series of subjects within the field of the history of elites. The intense contact with the western schools of thought allowed Romanian historians to widen their perspectives, and, as a result, to publish a series of works concerned with the different categories of the Romanian elite, especially for the period after 1867: the secular or ecclesiastical elite (Sigmirean 2000, 2007), the economic elite (Dobrescu 1996, 2006), or the political elite (Popovici 2010). As can be seen, many categories of the elite have not been discussed, among them the category of civil servants. Furthermore, while the abovementioned works are useful from the standpoint of their factual contributions, they lack a theoretical and methodological framework that could be applied to other elite categories.

The prosopographic study of the Romanian civil servants has also been neglected by the main authors who have dealt with the history of the liberal period (Retegan

1979, 2004; Suciu 2000). Renowned specialists in their field, the aforementioned authors created well-documented syntheses, with an obviously necessary general view, which encompassed several political actors. Their goal was to reconstruct the Transylvanian political scene in a positivist manner and to emphasize the power relations between the Romanians, Hungarians and Saxons, by using a politicalnational interpretation grid. They outlined the stages of administrative reorganization, the increase in number of the Romanian civil servants, as well as their presence in the different political decision-making structures of the Romanian national movement. However, the presence of conflicts arising from the failure of obtaining certain administrative offices has been omitted, despite the fact that it constitutes a part of the recruiting process of these civil servants. Furthermore, certain value judgments were made concerning the national activity of some and the lack of empathy of others with their nations' desiderata, without a proper emphasis of the dual character of their political-administrative position. Finally, a real evaluation of the role played by the Romanian civil servants in the designing of a Romanian political strategy was not provided.

Thus, the lack of studies strictly concerning the Romanian high civil servants in Transylvania can be easily noticed. Judit Pál's study (Pál 2008) comes closest to the intentions of the present paper, but her study is not concern only with the Romanians. The author's analysis of the Transylvanian Lord Lieutenants, the great majority of which were Hungarians, offers an extremely useful methodological framework, which can be applied, with few exceptions, to the analysis of the civil servant body during the liberal period.

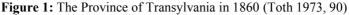
3 Data and Methodology

3.1 Preliminaries and Sources

First, we need to establish what Transylvania meant as an autonomous region in 1861. In December 1860, the former Great Principality of Transylvania lost some of its Western regions: Zarand and the Partium (Közep Szolnok and Kövár). As a consequence, the territorial composition of Transylvania that we take into consideration was the following: 8 counties (vármegye), 2 regions (vidék), 5 Szekler seats (szék) and the Saxon Settlement (Királyföld) (see below Figure 1).

Second, we need to circumscribe those categories of clerks that can be considered part of an elite layer, meaning the ones who occupied a position above the level of copyist (fogalmazó) at the Transylvanian Court Chancellery in Vienna and the Gubernium in Cluj. We shall also take into consideration the civil servants having the rank of assessor (ülnök) or higher at the High Court of Justice of Transylvania (Királyi Tábla) in Marosvásárhely/Târgu Mureş, the Appeal Court of Sibiu (Királyi Főbiztosság), and the later High Court of Justice (Királyi Legfőbb Tőrvényszék), founded in 1865. At local level, we are interested in the following high-ranking stations: Lord-Lieutenant, Captain-General or county administrator (főispán / főkapitány / főispáni helyettes), Vice-Lord-Lieutenant and Vice-Captain-General (alispánok / alkapitány), county notary public (főjegyző), and royal chief judge or

president of the county court of law (főbirák / törvényszék elnök). It should be stated that some of these previous functions meant, basically, the same thing, and their holders shared equal position (i.e. Lord-Lietenant and Captain-General), the difference being related to the old, late-medieval, administrative situation of the province.





As for our sources, the most important are the yearly-published directories of public employees, which are complete for this period of time ("Erdély Nagyfejedelemség Hivatali Tiszti Névtára"). Other relevant biographical information can be found in catalogues and encyclopaedias (mostly Diaconovich 1989-1904), but also in the Romanian journals from that period ("Telegraful Român", "Gazeta Transilvaniei"). Finally, various history works, mostly biographical (Suciu 2000), as well as editions of published documents (Hitchins, Maior 1970; Pascu et alii 1973-1987; Retegan, Suciu et alii 2006) complete the framework of our research.

Given the vastness of the subject, we have decided to dwell only upon some aspects of the history of Romanian civil clerks from Transylvania, so as to provide the essential elements for understanding its evolution during the studied period. Thus, the main topics we intend to approach in the current paper are: the social-professional and educational background of the Romanian bureaucratic elite, its recruitment procedures, the career evolution (*cursus honorum*) and, finally, the political involvement of these bureaucrats and their relation with the political activities of the Romanian national movement.

3.2 The social-professional and educational background of the Romanian high civil servants of Transylvania

Very few of the Romanian officials came from families with a political or administrative tradition which span the length of several generations. Most of them were descendents of school teachers or priests, and had benefited from scholarships awarded by the Greek-Catholic or the Orthodox Churches, thus being the first in their families to reach such high social status. There were, however, exceptions: the representatives of noble families, or of merchants or gold mines owners' families from the mountain area of Transylvania.

This Romanian "power elite" (according to Mills' definition – Mills 2000) was formed by members who had the same educational roots, graduated the same Colleges and knew each other from school time (sometimes earlier, due to kinship), thus showing a superior cohesion level after being appointed in offices. The Transylvanian society in its whole could be considered, at that time (1861-1867), a status-based society (as Weber presents it) so this knowledge of each other was very important for them, for their self and group image, in order to make them develop a class conscience and, further on, Meissel's three "C"s: conscience, conspiration, cohesion (Meisel 1962). Status was not only important to group cohesion and evolution, but also to individual development: from pecuniary aspects to marriage-related ones, status was essential for helping the members of the elite achieve "a good life" – in terms of material and social ideals.

We can assert that the emergence of this elite is a consequence of factors rooted back in the 18th century, the most important being: the educational one, the economic one, the mental and demographic ones. The first three factors we mentioned are strongly interconnected: as a consequence of the expansion of education among the Romanian population in the 18th century, both a mental environment and an economic progress were created, that led to the awareness of the need to create administrative and political elites. The fourth factor, the demographic one, merely provided a larger selection base for the first three, thus ensuring, through intrinsic competition, the entry into the elite ranks of the best representatives of the Romanian nation.

From a geographical point of view, no specific area in Transylvania can be considered a priority "extraction area" for the future bureaucratic elite. Its members came from all Romanian inhabited areas and only few of them got to know each other before attending the same schools. Nonetheless, it is obvious that, in some of the cases, the socio-economic circumstances have influenced certain areas. So it is the case of Braşov and its surroundings, an area heavily populated by Romanians, bordering Wallachia, thus benefiting of both the ecclesiastic and intellectual influence of the Romanians south of the Carpathians, as well as of the increased economic potential of all border regions. Ioan knight of Puşcariu, Ioan Alduleanu and Ioan Bran of Lemény, some of the most important characters of the liberal period, originate from this region. Another such area was Sibiu and its surroundings, due to the influence of the Orthodox Bishopric [latter Archbishopric]. The same stands for Blaj, the Greek-Catholic Archbishopric see. Cluj County (Cluj was the capital city) also brought up an important figure of that time: vice-governor Vasile Ladislau Pop. It can be stated that the proximity to cities, whether major centers (Cluj, Braşov, Sibiu) or boroughs (Abrud, Deva, Reghin, Blaj) contributed to the polarization of certain national energies and, in time, caused a stronger emergence of the bureaucratic elite members.

Regarding their studies, it is difficult to specify a single intellectual path for the entire category. For some, even the most basic biographical data is unavailable, for instance the year of birth. Either way, among the main secondary schools, one should count the following: the Greek-Catholic Gymnasium from Blaj, the Roman-Catholic Gymnasiums from Braşov and Sibiu, or the Royal Catholic Lyceum from Cluj. The lack of Romanian secondary schools (there was only one gymnasium, in Blaj) on the one hand and the multi-ethnic character of Transylvania on the other hand, determined the prospective elite members to attend educational institutions with Latin, Hungarian and German as teaching languages. This seemingly shortcoming must have become an advantage on the long term since, at a young age, they learned all the four languages used in the administration of the province. At the same time, access to secondary schools must have also been a screening filter for the less gifted pupils or for those whose parents did not have the needed financial possibilities. Last but not least, we can assess that secondary schools were the first stage of school crucible, the period during which the bureaucratic elite members-to-be got to know each other, to strike up friendships or enmities, to form their first impressions of one another. The most probatory case in this respect is that of the group of students from Blaj who got involved in the "Lemenian trial" (1843-1846) – a trial in which some of the students and teachers from the Greek-Catholic seminar in Blaj took a stand against the bishop Ioan Lemeni and several of his closest friends. Despite the fact that they lost the case and they were expelled or forced to retire, many of them will become after 1848, and latter during the liberal period, administrative clerks in Transvlvania, while maintaining the good relations they welded during their studies (Deteşan 2008).

A major characteristic of almost all Romanian civil servants is the completion of their studies at the Tabula Regia in Marosvásárhely, a genuine cradle for all high civil servants of Transvlvania irrespective of their ethnic origin. Some of them also acquired diplomas at some prestigious universities in Europe. Abroad studies did not necessary imply success for those who completed them, but there are at least three examples indicating that obtaining a degree at a prestigious university helped those who wanted to reach the higher bureaucracy. Iosif Hodos, Vice-Lord-Lieutenant of Zarand, obtained his PhD in law at Padova (Chiorean 1996), Dimitrie Moldovan, secretary of the Transylvanian Aulic Chancellery from Vienna, attended polytechnic studies in the capital city of the empire, where Vasile Ladislau Popp, vice-governor of Transylvania, completed his university studies. They are, however, exceptions, for most of the future Romanian clerks of the years 1861-1867 completed their education at provincial law academies. These institutions offered a diploma which subsequently granted them both the possibility to obtain a job in bureaucracy and the necessary relations to acquire the law practice certificate. This is the case of: Ioan Alduleanu, Ioan knight of Puşcariu, Nicolae Gaetan, Ioan Bran de Lemeny, Ilie Măcelariu (Suciu 2000; Josan 1997).

Therefore, it is obvious that, if after 1860, university studies in European centers became almost mandatory for those who wanted to join the elite ranks, officials trained between 1835 and 1855 did not feel this need, partly because they could advance on the social and professional ladder without university studies. In conclusion, we can affirm that the Romanian bureaucratic elite from Transylvania in the period 1861-1865 was, as a whole, at the lower limit of what could be called university training, the target being to graduate two years of law studies and to start the bureaucratic career as soon as possible, rather than to enrich the study resume.

3.3 The recruitment of the Romanian high civil servants of Transylvania

This subchapter aims to provide a succinct description of the recruiting mechanisms. It must be stated that, although during the entire liberal period the Romanian public opinion requested that all Romanian civil servants go through a national filter in order to ensure that they would represent Romanian interests at a bureaucratic level, it was not the Romanian nation which elected its administrative representatives. Instead, the Transylvanian Gubernium made the nominations, and the Aulic Chancellery in Vienna approved them in the name of the Emperor. Thus, the appointments were mostly directly made, without the approval of the nation from which the civil servants descended, in accordance with certain meritocratic principles or at least on the basis of a loyalty towards those who made the appointment. In some cases the proposals of the representatives of the Romanian nation were also considered, but in general the State was the one to choose its future employees.

In reaching a decision on the election of the civil servants, the overall political situation played an utterly important role. After a decade of military and militarized administration, the early liberal period was regarded as a favourable moment to revive the old practice of assigning a job, based on kinship and favoritism. This situation is perfectly emphasized by one of the Romanian higher civil servants, Ioan knight of Puşcariu, in his memoirs, when he recounts the explanation of Count György Haller regarding the refusal of the capitancy of Cetatea de Baltă County because it was not possible to satisfy "...about two hundred aspirants who were hungry for 20 years. It would have been woe to me, if I had not provided bread to all of them" (Puşcariu 1913, 84).

Furthermore, when the question of representing the Romanian nation at the administrative level arose, national affiliation became, in turn, a criterion for office appointment. Since December 21st 1860, an imperial rescript required to take "into respectful account the nations and denominations from the Great Principality of Transylvania, that formerly weren't represented or were not enough represented" ("Gazeta Transylvaniei" 1860, no. 61). In these circumstances, the only representative national institutions (the Orthodox and Greek-Catholic churches) felt the urge to become involved, through their hierarchs, in advancing proposals of civil servants. They tried to promote well-prepared people, able to defend the rights of the Romanian nation, while, at the same time, being loyal to their faith and the private interest of one or another of the churches. Hereabout, it should also be highlighted the political way in which the Romanian leaders understood the bureaucratic apparatus, this unclear perspective being perpetuated throughout the liberal periods (see below 3.5).

Notwithstanding, even within the nation, among denominations, there have been strong struggles for offices. At a collective level, the two religious leaders (who still adhered to a pre-modern type of thought) insisted upon the maintaining of a confessional balance within the Romanian civil servants. For example, when, in 1860, the Orthodoxes found out that the authorities wanted to appoint two Greek-Catholics as Aulic councilors, they attacked the one who was politically insecure (Gheorghe Anghel) in order to obtain his replacement with an Orthodox (Retegan, Suciu et alii 2006, 331-333). Then, in 1861, when the two denominations had to submit lists of proposals to the Viennese authorities in order to choose the future Romanian royalist MPs and civil servants, a flagrant imbalance between Othodoxes and Greek-Catholics was created. Albeit the Romanians tried to obey certain principles (priority to the higher offices for the senior civil servants, priority for those who made a radical display for the national cause), old habits such as nepotism and favoritism were also present: for example, the grandchildren of the Greek-Catholic Metropolitan (Iosif and Dionisie Șterca-Șuluțiu) were proposed for key positions, including that of Lord-Lieutenant.

It should be noted that, in addition to the lists of favourites, there were also lists of undesirable persons. Those Romanian officials who were considered not to have fought enough for the national cause were mentioned by name and their replacement was asked for. This was the case, among others, for Ioan Alduleanu, Paul Dunca, Alexandru Lazăr, Augustin Láday, Ladislau Hosszu and others, some of them with high ranks in the Transylvanian Gubernium or in the county administration. It is obvious that they preferred to plead in the spirit of the Austrian bureaucracy rather than in the Romanian National spirit (Retegan, Suciu et alii 2006, 583-588).

At an individual level, the competition became manifest through the exposing of certain pecuniary interests. For many Romanians, the civil servant office was the sole and best source of income to ensure an above-average standard of living. For this reason there have been cases where some candidates were subjected to personal denigration campaigns, orchestrated by their opponents. This was the case of Dimitrie Moldovan, who was about to become Aulic secretary at the Transylvanian Chancellery in Vienna, but was forced to face the powerful intrigues developed by Ioan Bran of Lemény. The latter turned for support to the Bishop A. Şaguna, who personally corresponded with the Austrian Ministry of Interior to impose his favourite. Eventually, the two endeavors failed, and D. Moldovan became Aulic secretary.

To summarize, between 1861 and 1867 there were two ways of access to the provincial bureaucracy in Transylvania for the Romanian clerks. The first one was their direct hiring by the authorities, and the second one was their positioning on the lists of proposals from the nation (represented by the Romanian churches). It is obvious that, in both cases, beyond their experience, skills and personal relationships played an important role, whether developed through the State or through the churches networks.

The former were certainly stronger and ensured a faster advancement, but sometimes involved neglecting the claims of the "nation", or of the church, and could result, ultimately, with the official label of "national renegade". The latter, the placement on the lists with national proposals, was also linked to good relationships with one or another of the churches and to the personal visibility inside the national movement. Besides, where, as previously asserted, most candidates had about the same education level, it was natural that favoritisms and personal influence became key factors in finding a bureaucratic post. As far as the Romanian clerks from the liberal period are regarded, we can not say that their family relationships have been able to create benefits beyond their political and national community. Nonetheless, it must be stated that, for the studied period, we have only found few cases where kinship or affinity relationships were a real advantage in the bureaucratic advancement (even the above-mentioned grandchildren of the Metropolitan Şuluțiu did not obtain the desired offices). Most of the Romanian civil servants were the first in their family to reach such social ranks and had not yet built all the necessary kinship networks. Hence, the professional integration was due to other types of channels (national radicalism, fidelity to one or another camp), not to kinship.

3.4 Patterns of administrative career within the Romanian high civil servants of Transylvania

Most of the Romanian officials in office in 1860 were people who took advantage of the political events of 1848-1849 and, especially, of the neo-absolutist period (1850-1860). They were granted administrative positions during a militarized regime, when the Hungarian political elite partly withdrew from administration or was purged since in 1848 had risen against the emperor. The positions they held in the years 1850-1860 were minor and personnel rotation was quite customary.

The perspective opened by the changes in 1860-1861 allowed those in office to advance and those outside the system to try to reach positions which granted further promotion. In terms of offices, those who reached the top hierarchy were few in number and, as previously mentioned, were among those whose bureaucratic experience was doubled, in many cases, by solid education. Among them, we can mention:

- Vasile Ladislau Popp, Aulic councilor (1861) and vice-president of the Transylvanian Gubernium (1862-1865),
- Dimitrie Moldovan, Aulic councilor (1862-1865),
- Alexandru Lazăr, councilor to the Transylvanian Gubernium (1861-1866),
- Petru Anca, Vice-Lord-Lieutenant of Dăbâca county (1861-1866),
- Mihail Orbonaş, assessor of the Imperial Court of Transylvania (1861-1866)

The inception of the civil servant's career varied, some officials reaching high positions quite quickly, while others remained to occupy less-important ones. Generally, those who reached offices in the central administration of the province seemed to be tempted to maintain their position, a matter perfectly explicable both by material advantages as well as by a more pleasant social environment in the provincial capital: the city of Cluj. Among those who reached the county and local administration there was a large fluctuation which can be explained both by the desire for advancement, and the geographical distance between the employment area and the stable residence. It was perfectly normal for a future clerk not to refuse a job in the state apparatus, even if it implied relocation to another county, and on the other hand, it was equally normal that at least some of those relocated to try to resettle to their

county of origin or to move into more favourable counties for themselves and their families. So far, there are no studies devoted to the geographical fluctuation of the civil servants, but the evolution of this process can be easily traced in the clerk's calendars of the time ("Erdély Nagyfejedelemség Hivatali Tiszti Névtára").

As far as the preservation of office is concerned, the situation was more complicated. Some of them were very longevous, as Alexandru Bohățiel who remained Captain-General of Năsăud from 1861 until 1875 (Buhățel 2009). Likewise, other lower-level officials (Ioan Pinciu) continued their career in the Hungarian state apparatus during the dualist period. Despite the fact that they were called "renegades" and the historiography has neglected their activity after 1867, their social position remained clearly superior to the majority of their conationals.

Others, however, preserved their offices for very short periods of time. Individuals like Ion Raţiu or Ioan P. Maior have chosen to make no political or national compromises, resigning quickly from their offices as Vice-Lord-Lieutenants of Turda County. They were among the few who have allowed themselves such a gesture, given that the first was already a successful lawyer in the area and the other came from a family of well-known Romanian traders. Ioan P. Maior was also the father-in-law of Vasile Ladislau Popp. Most of the civil servants avoided such extreme gestures, their own existence and their family's being in a close interdependence with the sinecure. In other cases (for example Ioan Alduleanu), the short periods they spent in various offices were due to constant advancement, the increased frequency of the bureaucratic positions representing, in fact, an advantage.

After concluding the 1867 Compromise, a large part of the Romanian officials were either changed or retired, being replaced by Hungarian ones. This process should not be regarded solely from a national perspective, but also from a social and political one. Politically, the new Hungarian clerks played key-roles in the elective process (Seton-Watson 1911), helping to maintain a favourable majority in Parlament for the deakist government. Socially, after half a decade of passivity, the financial situation of the Hungarian officials was of such nature that its reintegration into administration was a necessity. Nevertheless, there were also Romanian officials whose careers continued after 1867 (V.L. Popp, I. Pinciu). Their previous experience and their loyality to the principles of bureaucracy at the expense of national political involvement contributed to this fact.

3.5 The Romanian high civil servants and the Romanian national movement of Transylvania

As previously stated, this issue was most ardently debated in the Romanian historiography (Retegan 1979, 2004, Suciu 2000). The contributions of this study will therefore be limited to the general analysis of the degree of participation that these civil servants had in the main decision-making organisms of the Romanian nation: the national conferences from Sibiu (1861 and 1863) and the Permanent Committee elected at these events. At both conferences, an equal number of Greek-Catolics and Orthodoxes attended (50 in 1861 and 75 in 1863), some appointed by the hierarchs of the two denominations, some assigned at local level. In 1861, 10 (50%) out of the 20

members of the Permanent Committee were clerks, while in 1863, 18 (62%) out of the 29 members (Popovici 2010a). The growth between the two years represents a proof of the greater role the officials gained in the national movement, role that will increase with the elections for the 1863-1864 Sibiu Diet. Most Romanian MPs in this regional Parliament were clerks, this pointing out the interdependence between bureaucracy and politics in Transylvania between 1861 and 1867.

The civil servants' role was vital in politics, considering that they had access to the electoral lists and, according to these lists, the tactics for elections in each college were established. This issue is well captured by D. Moldovan in one of his letters to G. Bariţiu (1863): "Now, let's see how can we be reelected with a plurality. This is the way I evaluated and I also wrote to Axenti, Bishop Şuluţ, Fetti, and Pop in Sebeş and I am going to write to Bishop Şaguna, I then, ordered, my brother in Deva: 10 brave Romanians make up a committee, they win the whole catalog of contributions with even more than 8 fl.[the census for becoming elector], if not else with money. There we will see who is an entitled voter. Every fellow from those 10 wins other 10-20 proeminent Romanians. The latter win a brother, nephew, the father-in-law of his child, etc., etc." (Pascu et alii 1981-V, 155). As a result, the Romanian clerks got involved both in organizing the elections and in the representative activity itself as well.

We must state from the start that the involvement of the Romanian civil servants in the electoral campaign was not at all highly active: there were no electoral programs, no promises were made and no speeches were given. Out of this perspective, the campaign went on quietly. A whole lot of the candidates were elected without any problems, the candidacies set by the "inteligence" being validated by the voters. Generally, the elections of MPs for the Sibiu Diet lacked the violent convulsions this kind of events manifested in Hungary. No violence occurred, no murders or fights between rival factions. However, methods of electoral fraud were present in each electoral college. Starting with official forms (partial registering of the voters, retaining electors from reaching the polling centers, the erroneous recording of the vote, the incorrect counting of the voting options in favour of some candidates) and carrying on with unofficial forms (electoral bribery, intimidation) they spread all throughout Transvlvania, being widely reported by the press. The Romanian officials' role was to compensate, as much as possible, the Hungarians' eagerness to gain an overwhelming majority by illegal practices and to ensure the proper course of the electoral proceedings in the colleges where Romanians ran for office. They weren't successful everywhere, yet 48 Romanians MPs were elected in the Diet, out of whom 42 (70%) were members of the bureaucratic apparatus (Retegan 1979, 78 sqq).

During the Diet's sessions, these 42 civil servants have greatly contributed to the elaboration of the legislation recognizing the rights of the Romanian nation and the use of Romanian language in administration and justice. Their legislative expertise weighted heavily, proving to be essential in the fast elaboration of the new laws and in their voting.

After the Sibiu Diet was prorogued by the Emperor, in 1865, a new Transylvanian Diet was summoned in Cluj. In this new Diet, the number of Romanians was greatly reduced, most of them being regime loyal clerks appointed as "royalists" (*** 1866). The Diet in Cluj decided the annexation of Transylvania to Hungary, thus abolishing the autonomy of the province and anticipating the changes to come in 1867. Some of

the officials in the Sibiu Diet were later retired and became supporters of the Romanian passivity policy: not taking part in elections as a protest against the abolition of Transylania's autonomy.

It can be noted that, between 1861 and 1867, the civil servants were the best represented socio-professional cathegory, both in the internal structures of the Romanian national movement and also in the representative structures of the province. Their political role is evident and, given the premises, could not have been avoided: besides the clergy, the civil servants represented the most extensive elitist walk of life of the Romanian nation, being, naturally, its political representatives. However, this involvement would cost many of them after 1867, when the Hungarian authorities lost confidence in them.

4 Conclusions

This brief foray into the history of the Romanian civil servants from the 1861-1867 period binds us to stating some general conclusions on the subject.

It is obvious that, as far as the Romanian nation in Transylvania is concerned, the civil servants in the studied period represented one of the most important elitist cathegories, with great influence among their conationals, and also with great political authority. Socio-professionally speaking, the extraction stratum of most of the Romanian civil servants was the humble clergy and the rural intellectuals, which made them truly the first generation of elite. For the same reason, their professional path was less influenced by family relationships and more by the relationships they were able to build throughout their career, based on their fidelity and their competences. Few of the Romanian officials of the liberal period were university graduates, most of them having completed their education in Transylvanian law academies.

The recruitment of the civil servants was made directly by the Viennese authorities, yet there were notable attempts of the Romanian churches, as representatives of the nation, to impose certain favorites. The proposals of the Romanians were sometimes successful, but the final decisions were taken in Vienna. The internal struggles for the key positions were also present, since inter-denominational and inter-personal competition was quite strong.

Once they entered the bureaucratic system, the Romanian civil servants had to comply with two directions: one coming from the State which hired them, asking them to behave strictly bureaucratic, and the other coming from the churches which proposed them for office, expecting them to work in the interest of the nation. The stability of the civil servants seems to have been higher at the central level and lower below the county level. The carrer of many officials was interrupted by the implementation of dualism in 1867, yet there were also persons considered reliable by the new regime, whom thus kept their offices.

As far as the involvement of the clerks in the national movement is regarded, the statistical calculations have shown that they represent approximately 60-70% of the Romanian political elite, having, on the other hand, the advantage of an excellent law expertise. It should be emphasized, at this point, the duality of their political-national

position: as civil servants they were employed in an administration subordinated to the interests and the needs of the Monarch, though some of them took their quality as representatives of their nation in various administrative structures very seriously and did their best to gain their conationals an advantage. The apex of the political influence of the Romanian civil servants was represented by the 1863-1864 Diet. Subsequently, due to the political changes atop the Monarchy, the Romanian officials ended up losing all the gains of previous years.

As for future research paths, we stress the need to create a prosopography of the whole body of Romanian civil servants in Transylvania between 1861 and 1867, both for a better understanding of their political and social role, and because they represented the walk of life for the extraction of many of the future Romanian political elite members.

References

*** (1866) Protocolulu și cartea de documente a Dietei Marelui Principatu Transilvania conchiemata in libera cetate regia Clusiu pe 19-a Novembrie 1865 [The Protocol and the Documents of the Great Principality of Transylvania's Diet, Summoned in the Free City of Cluj on November the 19th 1865], Typography of Ioan Gaman, Cluj.

ACA-GCAF, Alba County Archives, The Greek-Catholic Archbishopric Fund.

Barițiu, G. (1889-1891) Părți alese din istoria Transilvaniei pre doue sute de ani din urmă. [Selected Parts from the History of Transylvania during the last Two Hundred years], volumes 1-3, W. Krafft Typography, Sibiu.

Buhățel, T. (2011) Din "Panteonul" marilor patrioți români ardeleni: Alexandru Bohățiel [From the "Pantheon" of the Great Romanian Patriots from Transylvania: Alexandru Bohățiel], Diotima Publishing Press, Cluj-Napoca.

Chiorean, I. (1996) *Iosif Hodoş. Viaţa şi activitatea sa [Joseph Hodoş. His Life and Activity]*, Mica Doris Publishing House, Târgu-Mureş.

Cosma, E. (2008) Dicționarul personalităților revoluției de la 1848 în Banat și Transilvania [Dictionary of Personalities of the 1848 Revolution in Banat and Transylvania], Romanian Academy Publishing Press, Bucharest.

Dăianu, E. (1925) Al doilea președinte al Asociațiunii: Vasile L. bar. Pop 1819-1875 [The Second President of the Association. Vasile L. baron of Pop], The Publishing House of the Association, Sibiu.

Deteşan, D. (2008) *Procesul Lemenian 1843-1846 [The Lemeny Trial 1843-1846]*, Cluj-Napoca University Press, Cluj-Napoca.

Diaconovich, C. (1898-1904) *Enciclopedia Română [The Romanian Encyclopedia]*, volumes 1-3, W. Krafft Print, Sibiu.

Dobrescu, V. (1996) Elita românească în lumea satului transilvan 1867-1918 [The Romanian Elite and the Transylvanian village], Petru Maior University Press, Târgu-Mureş.

Dobrescu, V. (2006) Funcții și funcționalități în sistemul de credit românesc din Transilvania până la Primul Război Mondial [Functions and Functionalities in the Romanian Credit System till the First World War], Petru Maior University Press, Târgu-Mureş.

Hitchins, K., Maior, L. (1970) Corespondența lui Ioan Rațiu cu George Barițiu 1861-1892 [The Correspondence between Ioan Rațiu and George Barițiu 1861-1892], Dacia, Cluj.

Josan, N. (1997), Ioan Puşcariu (1824-1912). Viaşa şi activitatea sa [John Puşcariu (1824-1912). His Life and Activity], The National Museum of the Union, Alba Iulia.

Netea, V. (1974) Lupta românilor din Transilvania pentru libertate națională 1848-1881 [The Struggle of the Romanians from Transylvania for National Liberty between 1848 and 1881], The Scientific Publishing House, Bucharest.

Meisel, J.H. (1962) *The Myth of the Ruling Class. Gaetano Mosca and the Elite,* University of Michingan Press, Ann Arbor.

Mills, C.W. (2000) The Power Elite, Oxford University Press, New York.

Pál, J. (2008) The Transylvanian Lord-Lieutenants after the Austro-Hungarian Compromise, in Karady V. and Török B.Z. (eds.), Cultural Dimensions of Elite Formation in Transylvania (1770–1950), EDRC Foundation, Cluj-Napoca. Pascu Şt. et alii (1973-1987) George Bariţ şi contemporanii săi [George Bariţ and his Contemporaries], volumes 1-8, Minerva Publishing House, Bucharest.

Pascu, Şt. (1978) Românii din Transilvania împotriva dualismului austro-ungar 1865-1900 [The Romanians from Transylvania against the Austrian-Hungarian Dualism 1865-1900], Dacia, Cluj-Napoca.

Păcățian, T.V. (1904, 1905) Cartea de aur sau luptele politice-naționale ale românilor de sub Coroana Ungariei [The Golden Book or the Romanian Political and National Struggle under the Hungarian Crown], volumes 2-3, Iosif Marschall Print, Sibiu.

Popovici, V. (2010) Acte și documente privind elita politică românească din Transilvania 1869-1896 [Papers and Documents regarding the Romanian Political Elite from Transylvania 1869-1896], Mega Publishing House, Cluj-Napoca. Popovici, V. (2010a) Elita politică românească din Transilvania (1861-1881). O perspectivă alternativă [The Romanian Political Elite from Transylvania (1869-1881). An Alternate Perspective], in Annales Universitatis Apulensis. Series Historica, 14/I, 213-227.

Puşcariu, I. (1913) Notițe despre întâmplările contemporane [Notes on the contemporary facts], The Archbishopric Typography, Sibiu.

Retegan, S. (1979) Dieta romănească de la Sibiu, 1863-1864 [The Romanian Diet from Sibiu, 1863-1864], Tribuna Publishing House, Cluj-Napoca.

Retegan, S. (2004) *Reconstrucția politică a Transilvaniei în anii 1861-1863 [The Political Reconstruction of Transylvania in the years 1861-1863]*, Cluj University Press, Cluj-Napoca.

Retegan, S., Suciu, D. et alii (2006) Documente privind mişcarea naţională a românilor din Transilvania [Documents Regarding the National Movement of the Romanians from Transylvania] volume 3, The Romanian Academy Publishing House, Bucharest.

Seton-Watson, R.W. (1911) Corruption and Reform in Hungary. A Study of Electoral Practice, Constable & Co. Ltd., London.

Sigmirean, C. (2000) Istoria formării intelectualității românești din Transilvania și Banat în epoca modernă [The History of the Intellectual Formation of Romanians from Transylvania and Banat], Cluj University Press, Cluj-Napoca.

Sigmirean, C. (2007) Intelectualitatea ecleziastică: preoții Blajului, 1806-1948 [Ecclesiastic intellectuals: the Priests from Blaj, 1806-1948], Petru Maior University Press, Târgu-Mureş.

Suciu, D. (2000) Antecedentele dualismului austro-ungar și mișcarea națională a romănilor din Transilvania [The Antecedents of the Austro-Hungarian Dualism and the Romanian National Movement in Transylvania], Albatros Publishing Press, București.

Toth, A. (1973) Parteien und Reichstagswahlen in Ungarn 1848-1892 [Political parties and Elections in Hungary 1848-1892], R. Oldenbourg Verlag, München.

Press

"Erdély Nagyfejedelemség Hivatali Tiszti Névtára"

"Telegraful Român"

"Gazeta Transilvaniei"

Political and diplomatic relations between Romania and the Ottoman Empire (1878-1900)

Cecilia Alina Sava, Ph.D. candidate1

Faculty of History and Philosophy, "Babeş-Bolyai" University 1 Mihail Kogălniceanu Street, Cluj-Napoca, 400084, Romania E-mails: cecilia.sava@ubbcluj.ro; alinutzasava@yahoo.com

Abstract. The Berlin Congress, which ended the Russo-Turkish war of 1877-1878, represented an international recognition of Romania's independence, even if some of the great powers agreed to this only provided that certain conditions were fulfilled. Turkey, who came out of this conflict deeply affected, was the first to recognize the sovereignty of its former vassal, without any pretence.

The present study aims to outline and discuss the evolution of the diplomatic and political relations between Romania and the Porte (which were re-established in the fall of 1878) during the last two decades of the 19th century. This subject has not been adequately discussed in historiography, while the secondary literature for this topic is lacking in details. Consequently, this study is the result of a large quantity of information synthesised and gathered from unpublished sources currently held by the Diplomatic Archives of the Ministry for Foreign Affairs from Bucharest.

In order to present the Romanian diplomatic thought and the strategies employed in the negotiations with the Turkish representatives, but also to objectively convey the spirit and the atmosphere in which the amicable relations between these two

Babeş-Bolyai University, Cluj-Napoca, Romania

¹ Investing in people!

Ph.D. scholarship, Project co-financed by the SECTORAL OPERATIONAL PROGRAM FOR HUMAN RESOURCES DEVELOPMENT 2007 - 2013

Priority Axis 1. "Education and training in support for growth and development of a knowledge based society"

Key area of intervention 1.5: Doctoral and post-doctoral programs in support of research. Contract nr.: **POSDRU/88/1.5/S/60185** – "INNOVATIVE DOCTORAL STUDIES IN A KNOWLEDGE BASED SOCIETY"

states were conducted in the abovementioned timeframe, we have analyzed (from both a quantitative and a qualitative perspective) the diplomatic reports, and the notes and telegrams sent or received by the Romanian plenipotentiary ministers at Constantinople.

Keywords: *Romania; independence; diplomacy; external policy; the Ottoman Empire.*

1 Introduction

The establishment of the Romanian Legation in Constantinople and the reestablishment of diplomatic relations between Romania and the Sublime Porte

In the "biography" of the Romanian state, sovereignty has been the fundamental characteristic, along with territoriality, centralization, bureaucracy and constitutionalism. By 1878, these later objectives had already been accomplished.

The Russo-Turkish war of 1877-1878 which followed the rebellion of 1875 in Bosnia and Herzegovina created a favourable situation for Romania, who had been eagerly awaiting the right moment to proclaim its independence and to obtain it through its own military efforts. Therefore, on the 9th /21st of May 1977, Mihail Kogălniceanu – the minister for foreign affairs -, solemnly proclaimed Romania's independence and declared the breaking of diplomatic relations with the Porte. The victories obtained by the Romanian armies at Plevna (1877) and Vidin (1878) encouraged both the political elite and the public opinion. However, Romania's political efforts did not come to a halt, as the state's sovereignty also had to be recognized by the Great European Powers, who initially appeared reserved on the matter.

At the San Stefano Congress and afterwards at Berlin, Romania was placed in an unexpected situation: the recognition of its independence seemed less straightforward than the political authorities in Bucharest had surmised – some of the Great Powers delayed this matter until 1880. Surprisingly, Turkey immediately and unconditionally recognized the independence of its former vassal, who, in turn, would prove to be a faithful ally and supporter of the Sublime Porte during the following decades of the 19th century.

Diplomatic relations between Romania and Porte were re-established immediately and ceremoniously, only three months after the end of the Congress which had created a new map of South-Eastern Europe. Following the unconditional and swift recognition of Romania's independence by Turkey, the reestablishment of diplomatic relations was somewhat expected. In September 1878, the exchange of diplomats between Romania and Turkey was carried out – D. Brătianu was appointed Plenipotentiary Minister and Extraordinary Envoy to Constantinople and in the autumn of the same year, the High Porte accredited his counterpart Suleiman Bey to Bucharest (11/23 October 1878), with the same title and rank. The moment of the Romanian diplomat's reception at the palace, as well as the reception banquet hosted in honour of his arrival at Constantinople were described by Brătianu in terms of their rare greatness. The sultan was profoundly impressed with the Romanian representative's praise and the salutations and wishes conveyed through him by the prince Charles. The Ottoman Emperor also expressed his wish to establish and maintain prosperous and friendly relations with Romania, to which he foretold a shining future.

The following diplomats successfully represented Romania at "His Majesty the Sultan's" court until 1900:

- Dimitrie Brătianu (1878-1882)
- Petre Mavrogheni (1882-1885)
- George M. Ghica (1885-1886)
- Ion Bălăceanu (1886-1890)
- Mihail Mitilineu (1890-1896)
- Trandafir G. Djuvara (1896-1899)
- Alexandru I. Ghica (1899-1902).

The general consulate in the capital city, as well as those in Thessaloniki, Bitola, Ioannina, Ruse, Cairo, Adana, Izmir and Teheran depended on the Romanian Legation in Constantinople.

Thus, immediately after the Berlin Treaty had been finalized, and Romania had gained a new international status, the diplomatic relations between this state and the Sublime Porte were re-established, but at a higher level than before, with more boldness and confidence on the part of the Romanian state. These rapports would prove to be extremely prolific, even in the troubled times that were to come, especially for the Ottoman Empire.

2 Literature Review

The study of Turkish-Romanian relations has not been given ample attention in historiography, and as a consequence, the secondary literature concerning the evolution of the rapports between the newly independent state and its former suzerain is severely lacking in information.

This paper is based on two categories of sources: unpublished archival documents and published sources. The latter are represented by memories and relevant secondary literature, despite its previously-mentioned scarcity. The combined use of these types of sources can offer the premises for drawing some solid conclusions regarding the bilateral relations between the two states, at the end of the 19th century.

From the point of view of historical research, the most valuable sources are those that remain unpublished, that are found in the Diplomatic Archives of the Ministry of Foreign Affairs. It is here that we have encountered several archival funds which offer detailed information regarding this subject. Among these, the most interesting, extensive and well-structured one is the Constantinople Fund, which offers information concerning the establishment of the Romanian Legation in Constantinople, its functioning (staff, administrative structures), as well as that of the Romanian consulates on Ottoman territory. The recognition of Romania's independence by Turkey, the political relations between the two countries during the last two decades of the 19th century, and especially the aid given to the Porte by Romania in its conflicts with various Balkan states deserve to be researched and can be reconstituted on the basis of the precious data found within this fund.

On the matter of political relations, Fund 21 has been found very useful from the perspective of the Sublime Porte's opinions on Romania in the last decades of the 19th century.

While the list of the Romanian diplomats who were accredited to Constantinople, as well as that of the Romanian consuls on Ottoman territory can be reconstituted, the same cannot be done for the Sublime Porte's envoys to Bucharest. Their full names, the order in which they occupied the position of plenipotentiary ministers in Bucharest, and even their activity in Romania's capital are issues which are currently absent from the segment of diplomatic relations.

The general works mentioned at the end of this study have greatly contributed to the clarification of the international context within which the bilateral Romanian-Turkish relations were established, as well as to that of the internal situation in each of the two states, a factor which in many cases influenced the diplomatic strategies.

3 Proposed Methodology

The information extracted from telegrams, notes and addresses found in the archives, as well as that from published sources and general literature was thoroughly analyzed, using a method of intersection research by combining a qualitative and a quantitative approach, in order to accurately and objectively present the nature of the diplomatic and political relations between the two states as it can be deduced from the primary sources.

4 Analysis

4.1 European context of Romanian-Ottoman relations and the rapports with other states

After diplomatic relations with Turkey were re-established, these evolved cordially during the final two decades of the 19th century, a period when Germany dominated the European political scene. The quasi-hegemony of this empire was perpetuated by the "Iron chancellor", who created, maintained and dissolved alliances and systems of alliances until 1890. In order to neutralise France, the German chancellor brought to his side Austro-Hungary and Italy, but also smaller states, such as Romania, who wanted to establish itself as a centre of South-Eastern European stability, and to ensure a minimum of security against Russia.

Around 1890, after Bismarck's resignation, the rapports between the Great Powers were extremely complex. France and Germany could not come to terms due to the

issue of Alsace and Lorene; the French republic and England were caught up in their wide-ranging colonial politics, while Italy was waging a veritable customs war. Furthermore, Russia was on the verge of entering a war with England due to the areas of influence in Central Asia, and was closely following Austro-Hungary's interests in the Balkan Peninsula. To this were added the U.S's as well as Japan's – a state who was gradually establishing itself – tendencies to assert themselves as great powers.

However, at the end of the 19th and the beginning of the 20th centuries, France, Russia, and Great Britain signed a series of bilateral agreements which ultimately led to the Triple Agreement. The purpose of this alliance was to counteract the pretences of the Triple Alliance states.

Romania constantly manifested a friendly policy towards its neighbouring states in the Balkan Peninsula, a policy which was meant to preserve the status-quo. With the exception of the breaking of diplomatic relations with Greece between 1892 and 1896, due to the succession of Vanghelis Zappa's fortune, Romania tried to set an example for the Balkan states. Immediately after the 1878 moment, these states developed and implemented a policy of expansion, which was encouraged by the political and economic crisis that Turkey was undergoing. Particularly, Greece, Serbia and Bulgaria were concerned with the liberation of certain territories which were under Ottoman domination, and intended the subsequent integration of these territories within their own borders. Romania's own goals were comparatively more difficult to attain, in the sense that Bessarabia and Transylvania, the main targets of its European policy, were under the dominion of two great empires, Russia and Austro-Hungary, two states that were not showing any weaknesses or signs of frailty. Consequently, the kingdom north of the Danube directed its attention towards regions of minor interest, as for instance the protection of Dobrudja from the encroachments of the neighbouring autonomous Principality or the fate of the Aromanian population that inhabited the Ottoman Empire.

Even if they were aware of the impossibility of claiming a part of the Ottoman province inhabited by the Aromanians, the authorities in Bucharest attempted, on the one hand, to improve the social, cultural and religious life of their co-nationals and, on the other hand, to maintain the friendliest and sincerest relations with the Ottoman Empire, which was in a state of visible decline. In this way, Romania hoped to obstruct, or at least to control Bulgaria's interests in the area, but also in Dobrudja.

4.2 The evolution of the political and diplomatic relations between the Romanian state and Turkey

A short while after South-Eastern Europe settled its new borders according to the Treaty of Berlin, a new military conflict would mark the Balkans, namely that between Serbians and Bulgarians which occurred in the spring of 1885. Romania would only have a mediator's role in this issue, due to the prestige it had gained with the recognition of its independence, and also owing to the negotiations it was presently conducting with the Patriarchy of Constantinople for the recognition of the autocephaly of the Romanian Orthodox Church. In a report coming from Istanbul it was stated that at a reception, Said Pasha had expressed his satisfaction regarding the dignified and appropriate attitude that the Romanian state had had during the latest

events. At the peace conference in Bucharest which was organized with the purpose of settling the differences between the two belligerent states, the Turkish delegate stated that he was glad to see peace restored to the Balkans and wished to thank the representatives of the Romanian Kingdom for their role in resolving this matter.

Another important moment in the Turkish-Romanian relations was constituted by the year 1897, a year which witnessed the outburst of the war between Turkey and Greece. In these circumstances, the Romanian representative to the Porte, Tr. Djuvara, along with the minister of justice, G. Djuvara, were granted an audience with the sultan on the 18th/30th of April. The representative of the Romanian government evoked the countless proofs of friendship that Romania had shown Turkey after 1878, and requested the following: that the Romanians in the Ottoman Empire should be permitted to renounce the Greek Patriarchy and to unite with the Holy Synod of the autocephalous Romanian Church, the recognition of Bishop Antim, that the Romanian government enjoy the regime of capitulations in the Empire, and the signing of a convention to extradite and resolve the problem of the right to property for the Romanians living in Turkey. In exchange for these requests the Romanian government notified the sultan that they were "so disposed as to allow the founding of a mosque in Bucharest and to contribute to its building". Furthermore, they promised to strive to improve the life of the Muslims in Dobrudja.

The War between Turkey and Greece and the assembly of troops by the Serbian and Bulgarian Governments on its borders put the Ottoman Empire in a difficult situation. Consequently, Turkey insisted on a defensive treaty with Romania. This alliance would have eased the Empire's situation, because Romania enjoyed a great political and moral authority in South-Eastern Europe.

The padishah's ardent wish to gather the Balkan states, including Romania, in a federation that would hold against the attempts at conquering the Empire by foreign states remained however unaccomplished. King Charles stated, in a letter addressed to Kiazim Bey, the Turkish minister, that the first condition for the signing of a treaty of alliance was to ensure its secrecy; this first condition had not been met. Indiscrete commentaries had been made in newspaper articles. All the rumours regarding this issue should be silenced, in order to be able to rekindle the attempts, later on, when the time was right.

Thus, the treaty of alliance could not be signed because Turkey did not respond favourably to Romania's request of recognition for the Bishop Antim. While the Aromanians' complaints were disregarded, the Bulgarians in the Ottoman Empire were assured that they would receive another 4 or 5 berats for new bishops. The Grand Vizier justified this action to the Romanian minister at Constantinople, T. Djuvara, by asserting that the appointment of new bishops in an already extant Church differed greatly from the establishment of new Church within the state. However, if Romania agreed to sign a treaty of alliance with Turkey, he hoped that the issue of the Aromanian bishop's recognition would be resolved. Djuvara considered that, within the late 19th century political context, a treaty of alliance with Turkey would 'drop as a bomb out of the blue', and that the Great Powers would regard the Romanian state with suspicion afterwards. The purpose of this alliance would have been the strengthening of Turkey's European territories, the Grand Vizier asserting that '... Bulgaria and Serbia would be insignificant if Romania sided with Turkey.'

Despite this, the great Turkish newspapers published many articles in praise of Romania's correct attitude towards Turkey. The fact that the king should receive the title of protector of peace and tranquillity in the Balkans was underlined. For instance, in Malumat, on May 20th 1897, an article entitled La Turquie et la Roumanie appeared. It began by praising the Romanian Government because "it has always had a conciliator, pacifying, amicable and correct attitude towards the government of the Ottoman Empire, and this attitude has been maintained during a time of difficult situations and political complications for the Porte. Resigned, as well, in the Hellenic question, the Bucharest Cabinet, faithful in its behaviour, has declared its solidarity with Turkey. (...) Romania, who had interests regarding other Balkan states, postponed them and supported the Porte. Turkey can only learn from and highly appreciate the royal Government's attitude and express its gratitude. (...) Romania, through the relations it entertains with the great powers, represents a model of peace in Europe." The Ottoman authorities were thus convinced that the Romanian state would not change its attitude towards the Empire in the future and would continue to aid it. Turkey thanked Romania for its gesture, even if the latter had not previously signed an agreement with the High Porte.

The article published in the same newspaper, on the 4th/16th of August 1897, significantly entitled *Le Gouvernement Impérial Ottoman et Sa Majesté le Roi Charles I er de Roumanie*, represented the finest expression of Romania's superior position in South-Eastern Europe: "For the balance in the Balkans and the progress of the Balkan states, Romania plays the most important part due to its prosperity and progress, which has been achieved with an admirable perseverance. King Charles I has always opted to maintain peace, which he considers necessary for development. For him, peace leads to progress in art, science and commerce. [...] With the approval of all neighbouring states, the King should receive the title of protector and conserver of peace and quietude. [...] The King, who has received our sympathy and the profound honour of all Romanians, has managed, as well, to gain the sincere affection of the Muslims in Dobrudja".

5 Conclusions

As it has been seen, the political and diplomatic relations between the two countries evolved favourably for both parties. Romania's gain was expressed through the primary role that it earned in the Balkan Peninsula, but also through the improvement of its relations with the other European powers. The prestige that the state had gained emboldened its neighbours, who would gradually remove the influences and meddling of suzerain powers in this area as well as those of the Western countries, which also had strong interests in the Balkan region.

Turkey also had nothing to lose from maintaining amicable relations with the Romanian state. Romania's military and economic development in the 1880s resulted in its transformation into a reliable support for the Ottoman Empire. This support proved to be indispensable for the Porte, especially in regulating its relations with the newly sovereign Balkan states, but also with the nations that strived for independence.

6 References

Archives:

Arhiva Diplomatică a Ministerului Afacerilor Externe [The Ministry of Foreign Affairs Diplomatic Archive]:

- Constantinople Fund, Personal Section, vol. 26/1878-1922; vol. 40-44/1869-1897

- Constantinople Fund, Political Section, vol. 132/1878-1920

- Fund 21/1878-1913, vol. 34/1897; vol. 36/1898

- Historic Archive Fund, Constantinople, vol. 304/1880-1884

Ekrem, A M (1994) Din istoria turcilor dobrogeni [From the history of the Dobrudjan Turks], Kriterion, Bucharest

Gallagher, T (2001) Outcast Europe: the Balkans, 1789-1989, from the Ottomans to Miloševič, Routledge, London; New York

Jelavich, B (1984) *History of the Balkans*, Cambridge University Press, Cambridge, London, New-York

(2004) Organizarea instituțională a Ministerului Afacerilor Externe. Acte și documente [The institutional organization of the Ministry of External Affairs. Deeds and documents], vol. I (1859-1919), Fundația Europeană Titulescu, Bucharest

Quataert, D (2005) The Ottoman Empire, 1700-1922, CUP

Şükrü Hanioğlu, M (2008) A brief history of the late Ottoman Empire, Princeton University Press, Princeton

Zurcher, E J (2004) Turkey: A Modern History, Revised Edition, I.B. Tauris.

Governmental and non-governmental action against organized crime and drug trafficking in Serbia

Darja M. Koturovic¹,

¹ University of Sheffield, UK; South East European Research Centre, Greece

dkoturovic@seerc.org

Abstract. Contemporary conceptualizations support that organised crime groups operate as loosely-organized networks of cells with less obvious chains of command interrelated through diverse transnational networks. The boundaries of conventional international cooperation in the field of law enforcement and judiciary have changed, mainly due to swift mobility, significant diversity of structure, activities and communication methods of criminal networks. Studies indicate that Serbia represents major transit country within the 'Balkan route', the main transnational path used to smuggle drugs into Europe. Ongoing reform process is aimed at demonstrating that Serbia will act as a security partner capable of preventing spreading of crime into Europe, but the effects of state actions have not yet been assessed. The study explores the role of state institutions in suppressing drug trafficking and attempts to identify potential structural holes that hinder the effectiveness of such efforts. Social network analysis is deployed in an innovative manner to map the state institutions' network designed to address drug trafficking groups. Potential identified obstacles may put forward the importance of redesigning the system on the basis of international cooperation and lessons learned.

Keywords: drug trafficking, organized crime, social network analysis, Balkan route, corruption

1 Introduction

Contemporary conceptualizations support that organised crime groups operate as loosely-organized networks of cells with less obvious chains of command and numerous affiliations interrelated through diverse transnational networks (Finckenauer, 2005; Morseli, 2009; Paoli & Fijanoau, 2004; Shelley, 2005; Williams, 2001). The boundaries of conventional international cooperation in the field of law enforcement and judiciary have changed, mainly due to swift mobility, significant diversity of structure, activities and communication methods of criminal networks. Reports indicate that South East European states lay on the Balkans route, the main transnational path used to smuggle drugs into European states (Europol, 2009;

UNODC, 2011). It appears that the geographical location, trade liberalization, and relatively low risk of law enforcement activities make this path pretty convenient. In this line, the fight against organized crime and drug trafficking remains one of the most significant priorities of the regional cooperation in SEE due to a direct threat to the interests of the region as well as the broader international community. Through numerous legislative changes and suppression activities, Serbian state attempts to send a message that it will act as a security partner capable of preventing spreading of crime into Europe.

2. Literature Review

The traditional perception of hierarchically structured criminal groups has been challenged, as contemporary research on organized crime characterizes many organizations as inherently complex, business-like structures that operate as loosely-organized networks of cells (Shelley, 2005; von Lampe, 2006; Williams, 2001). Consistent with these standpoints, social network analysis (SNA) is mainly utilized as an approach to study 'dark networks' (Garay Salamanca, Salcedo-Albarán, & de León-Beltrán, 2010; Williams, 2001).

SNA allows the exploration of the associations between interdependent individuals in social and geographic space and appears particularly useful in location or region-related investigations. Several concepts proposed in SNA may be utilized to facilitate the analysis i.e. 'nodes', 'cliques', 'cells', etc. (Hanneman & Riddle, 2005; Wasserman & Faust, 1994). These concepts, coupled with specific measures such as betwenness, density, centrality, provide the capacity to study inherently hidden networks and gain deeper understanding on the methods of their functioning (Antonopoulos, 2008; Morseli & Giguere, 2006)

Studies on criminal networks report a wide range of common characteristics of organized crime groups, regardless of the specific type of criminal activity performed. SNA enables evidence-based understanding of the network structure, provision of richer data on numerous individual interactions and more precise role positioning (Morseli, 2009; Klerks, 2001; Shelley, 2005). For instance, a study by Bruinsmai & Bernasco (2004) examined the market of heroin in Netherlands and reported that over 85% of heroin comes through the Balkan route. They also found that networks involved in large scale heroin trade are homogeneous, characterized by affective bonds among the members, high density and strong cohesion (Bruinsmai & Bernasco, 2004).

Several specific traits of criminal groups associated with recent past and socio-political conditions in SEE states have been recognized. Research indicates that organized crime groups in the region form an integral part of the worldwide "Crime Syndicates" and control most of the international narcotics trade (Trifunovic, 2007; UNODC, 2008). They are motivated by profit as a common goal rather than national interests, which allows close cooperation among diverse individuals regardless of their ethnicity or political beliefs. Additionally, the utilised routes of smuggling cigarettes, gas, oil, cars and other products throughout the region resulted in a strong traffic structure unaffected by state actions.

Research confirms that fluid network structure provides drug trafficking groups with numerous advantages including adaptability, wide-scale recruitment, resilience, as well as capacity for quick learning and innovation (Ayling, 2009; Shelley, 2005;). Moreover, networks have greater capacity to exploit new modes of communication and international collaboration than state actors whose activities are based on hierarchical models (Williams, 2001). Elusive quality of networks comes from their ability to increase market efficiency, coexist within hierarchical structures or without them, and be modelled in different ways (Williams, 2001). This type of organization provides greater efficiency, organizational flexibility and hinders law enforcement efforts to identify and position group members (Shelley & Picarelli, 2002). Finally, transnational criminal groups lack specific ideology, and often use political corruption as a tool for their ends, infiltrating in this way into the political system (Morselli & Giguere, 2006; Shelley, 2005).

3. Methodology

This research is unique in terms of utilization of social network analysis for examination of governmental actions in suppressing drug trafficking networks. Commonly applied for mapping criminal networks, this approach is deployed in an innovative manner to map the state institutions' network designed to address drug trafficking groups. Empirical evidence supports that presence of legitimate actors in a variety of criminal contexts enables resilience to organized crime groups and allows for profound interconnectedness with the system. In this context, the exploration of a potential symbiotic relationship among actors from upper-and under worlds will be performed by indicating specific nodes (individuals or institutions) that may represent important structural holes hampering state efforts.

The aim of this study is to investigate the actions undertaken by the Government of Serbia in combating organized crime and drug trafficking, as well as the potential obstacles of these actions related to the role of legitimate actors in criminal networks operating in the region. The study explores the role of state institutions in suppressing drug trafficking and attempts to identify potential structural holes that hinder the effectiveness of such efforts. Concurrently, the sphere of actions undertaken by the civil society in regards to drug trafficking and its impact in the society is investigated. Within this framework, the main research questions include the following:

- What is the role of the Balkan route as an international grid of illicit drug traffic?
- To what extent and in which way the network of state institutions addresses drug trafficking?
- What are the potential links among Serbian legal and illegal actors in the context of drug trafficking and what is the strength of these links?
- What is the role of Serbian CSOs in contributing to systematic documentation and study of this phenomenon?
- What are the implications for policy recommendations related to international cooperation in this matter and regional development?

The study will draw upon a variety of official sources, including court case management statistics; trial reports from Special Court on Organized Crime and Appellate court in drug trafficking and organized crime cases; reports of international organizations; annual CSOs' reports on drug trade, organized crime and money laundering; and important journal and newspapers articles by prominent authors in printed and online media. These multiple sources will be analyzed in order to map the existing knowledge on drug trafficking in Serbia and the results of state action in this regard.

The sample will include government officials from the main institutions in the field of combating organized crime and drug trafficking, as well as relevant CSO's representatives. Sample size will include approximately 60 to 100 participants. The potential participants have been identified primarily on the basis of their position. However, it is expected that some participants may indicate other important 'actors' who have not been previously identified (snowball effect). This technique can be a complementary strategy to obtain more comprehensive data on a specific issue. Elite interviewing will be utilized for data collection. Interviews will be semi-structured, formulated on the basis of the main research questions and official data from the institutions and international law enforcement agencies.

Certain constraints refer primarily to the research area of organized crime which is *per se* difficult to explore, as there is a lack of reliable data and sufficient reluctance from the official institutions to share valuable information. Further, the criminal networks are by their nature secret entities; therefore any information about their functioning must be obtained only once they are identified. That is, the available information refers only to those cases that have been traced or reported to the relevant law enforcement agencies. There are certain challenges inherent to the criminal network analysis such as information overload, high search complexity and profound reliance on the field knowledge. Finally, there will be an almost certain delay in accessing criminal justice data, common for the state administration and the judiciary. Nevertheless, the awareness of such possibility increases the chances to utilize the 'time gap' by collecting other data meanwhile.

4. Conclusions

The proposed research is expected to indicate certain political and social difficulties in Serbia, which hinder effective state-to-state cooperation in combating organized crime. Even though the Government of Serbia has made the fight against drug trafficking groups and corruption top priorities, the effects of state actions have not yet been assessed. Social network analysis appears to be the most appropriate method to analyse drug trafficking networks. Comprehending the key characteristics of criminal networks may assist in their identification and mapping. However, an indepth understanding of their emergence and modes of networking represents is a precondition to adequately address their impact on the society. Lessons learned show that weakness of the state and the extent of institutionalized corruption generate

greater vulnerability of the political system to withhold challenges. Acknowledging that political corruption and state capture by criminal networks are common features of countries affected by high levels of organised crime, it is of concern to identify important legitimate actors' positions as interest protectors with a 'brokerage capital'. In this context, criminal network analysis may offer important data regarding existing structural holes that have high potential to provide smooth functioning of illicit networks. This may put forward the importance of redesigning the system on the basis of international cooperation and lessons learned.

References

Antonopoulos, G. A. (2007). Cigarette smugglers: A note on four 'unusual

suspects", Global Crime, 8 (4), pp.393-398.

Ayling, J. (2009). .Resilience in Criminal Networks, Regulatory Institutions

Network, ANU, Centre for Excellence in Policing and Security.

Bruinsmai & Bernasco (2004). Criminal groups and transnational illegal markets:

A more detailed examination on the basis of Social Network Theory, Crime, Law &

Social Change, 41: 79–94, Kluwer Academic Publishers.

Europol (2009). European Union Organised Crime Report. Luxembourg: Office

for the Official Publications of the European Communities.

Finckenauer, J.O. (2005). Problems of Definition: What is Organized Crime?

Trends in Organized Crime, 8(3), 63-83.

Garay Salamanca, L.J., Salcedo-Albarán, E., & de León-Beltrán, I. (2010). Illicit Networks Reconfiguring States, Social Network Analysis of Colombian and Mexican Cases. Bogotá, Colombia, Metodo Foundation.

Hanneman, R.A., & Riddle, M. (2005). Introduction to social network methods. Riverside, CA: University of California, Riverside http://faculty.ucr.edu/~hanneman/. Klerks, P. (2001). The network paradigm applied to criminal organizations: Theoretical nitpicking or a relevant doctrine for investigators? *Recent developments in the Netherlands. Connections*, 24, 53–65.

Morselli, C. & Giguere, C. (2006). Legitimate strengths in criminal networks.

Crime, Law & Social Change, 45, 185–200.

Morselli, C. (2009). Inside criminal networks. New York: Springer.

- Paoli, L. & Fijnaut, C. (2004). *General Introduction*. In: C. Fijnaut and L. Paoli (eds), 1 -19. Phantoms of Organised Crime, Corruption and Terrorism, Nijmegen: Wolf Legal Publishers.
- Shelley, L. I. & Picarelli, J. T. (2002). Methods Not Motives: Implications of the Convergence of International Organized Crime and Terrorism. *Police Practice and Research*, 3(4): 305-318.
- Shelley, L. I. (2005). The Unholy Trinity: Transnational Crime, Corruption, and Terrorism. Brown Journal of World Affairs, 11(2): 101-111.

Trifunovic, D. (2007). Special Report: Terrorism and Organized Crime in South-Eastern Europe: The Case of Bosnia-Herzegovina, Sanzak, and Kosovo. Defence & Foreign Affairs/International Strategic Studies Association.

UN Office on Drugs and Crime (2008). Crime and its impact on the Balkans and affected countries. Vienna International Centre. <u>www.unodc.com</u>

UNODC World Drug Report (2011).United Nations Publication, Sales No. E.11.XI.10.

Von Lampe, K. (2006). The cigarette black market in Germany and in the United Kingdom. *Journal of Financial Crime*, *13*, 235-54.

Walker, B., Holling, C.S., Carpenter, S.R., & Kinzig, A. (2004). Resilience,
adaptability and transformability in social–ecological systems. *Ecology and Society 9(2):* 5. Retrieved from

http://www.ecologyandsociety.org/vol9/iss2/art5/

- Wasserman, S. & Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge, UK.
- Williams, P (2001) 'Transnational Organised Crime' in Arquila, J and Ronfeldt, D (eds.) Networks and Netwars: The future of terror, crime and militancy.
 www.rand.orgj

PARTY STRUCTURE AND ORGANIZATION IN POST-COMMUNIST ALBANIA

Explain Centre-Right Party Success in Albania

Anjeza Xhaferaj PhD candidate

European University of Tirana, Department of International Relations and Political Science

axhaferaj@uet.edu.al

Abstract. This study is an attempt to analyze the factors that have contributed towards the success of the Democratic Party of Albania. It seeks to introduce new dimensions in measuring the success of a political party. The study rejects the elements of 'office-holding' and 'policy – or performance-based' measures of success, because they are determined by a complex array of economic, political, social and institutional factors. The concept of "party success" is defined in terms of: 'breadth' and 'durability'. The approach used in this study has regarded the political parties as the real political actors that as such should assume the responsibility for the perceived failures. Neither the lack of political culture, nor the communist system, are the main factors to shape the activity and development of political parties in Albania. Both macro-institutional and historical structural approach has limited explanatory power for the success of the Democratic Party. Its success is a direct consequence of the decisions that the leaders of the political parties have taken. The research showed that the Democratic Party of Albania has been successful because and when its elite have been able to craft inclusive ideological programs and because of its extensive penetration in the country, which it owes to its structure as a mass party. The party has established patronage linkages with its sympathizers, the rank and file and militants and consequently has ensured their loyalty and maintained an extensive electoral penetration in country.

Keywords: *political party, electoral system, ideology, critical juncture, post-communism.*

1 Introduction

The aim of this study is to understand and explain the factors that contribute to the success of the centre-right in Albania. It seeks to contribute to the limited literature on the political parties in Albania and on another level on the limited literature on centre-right formations in Central and Eastern Europe.

The theoretical framework used as a starting point the existing theoretical frameworks that provide a critical understanding of the political parties in Central and Eastern Europe. These existing frameworks are: macro-institutional framework; historical structural framework and elite cohesion and ideological structure framework.

The hypothesis of the study is that the factors that explain the success of the centreright formations in Central and Eastern Europe are elite cohesion and ideological crafting. The main findings and conclusions confirms that the Democratic Party of Albania has been successful because its elite have been able to craft inclusive ideological programs and because the relations of the party with its base are based on patronage.

2. Existing Framework Analysis on Party Development

The current existing theoretical frameworks in the study of political parties are the macro-institutional approach, the historical structural approach and elite cohesion and ideological crafting. In the following section is given a short explanation of these approaches.

2.1 The Macro-Institutional Approach – Electoral systems

The study of the electoral system is important because it is considered as one of main institutional choices that highly influences and has fundamental consequences on the political system (Norris, 2004, pp.4). Its function is to provide the mechanism for selecting the legislators, the executive and local authorities.

Duverger argues that 'the simple-majority single-ballot system favors the two-party system' (1954, pp.217). Psychological and mechanical mechanisms eliminate the small parties. On the other hand, the same factors (mechanical and psychological) which cause the elimination of the weakest parties, in the proportional system function in the opposite direction by establishing multi-partism and putting an end to the tendency towards a two-party system (pp.248). There is plethora of research studies on the effect of the electoral system on the Albanian political parties. The trend is to consider the political system as a two-party one (Kajsiu, 2004, pp.18). The main reason for this is considered to be the majoritarian system (Kajsiu, 2004, pp.21; Cili, 2002, pp.9), even though the Albanian electoral system during the period 1992-2005 has been mix, a combination of majoritarian and proportional. The logic used to reach to this conclusion is that of Duverger law, even though Kajsiu adds that there are other factors, such as falling turnout in local and national elections (Kajsiu, 2004, pp.22). Those who don't vote are in general people who are unhappy with either of the two main political parties. Consequently, those who vote are in a great part

sympathizers and militants of both parties, this resulting in a perpetuation of the twoparty system. Furthermore, the majoritarian system is blamed for empowering authoritarianism and undermining alternative leadership within the party because 'in a situation of a two-party system, the supporters of the party do not shift preferences to dissident groups: they are characterized by loyalty towards the party' (Cili, 2002, pp.10).

The current research on the Albanian elections has put more emphasis on the system rather than on agency. However, Sartori argues that 'the party arrangements and electoral systems only express the deeper determinants of the society' (1966, pp.167), thus placing more emphasis on agency rather than structure. Therefore, the electoral system rather than a cause is a product of an array of political and social factors that constitute the society. In Sartori words, the electoral system is 'the most specific manipulative instrument of politics' (1968, pp.273) and 'a two-party system can hardly be imposed from above merely through electoral devices' (1966, pp.175). As such it can be changed to reflect the needs of those how designed it. The choice of manipulation is wide indeed. It varies from the type of the system such as proportional, majoritarian or a mix of them to the specific rules for each system. Even the much praised proportional system does not necessarily produce a multiparty system. The outcome depends a lot on the size of the representative body, the district magnitude, the threshold (Liphart, 1994, pp.10-12; Shugart and Wattenberg, 2001, pp.10). The outcomes of the elections will depend a lot on the rules and rules are determined by the political actors to protect their interests.

In conclusion it could be said that the electoral system rather than a shaper of the political parties and party system is a product of them. It has reflected indeed the observation of Diamond and Platneer that 'among the many structural and historical variables that affect democracy, few are more open to rapid and intentionally designed change that the electoral system' (2006, pp.ix). Elections, after all are political. Therefore, I conclude that the electoral system cannot give an explanation for the success of the Democratic Party of Albania.

2.2 Historical Structural Explanations

Communist legacies and path dependence theories are often considered as being in the same vein, and thus scholars tend to study them as a single approach (Clark, 2002, pp.28). However, considering those as different and deserving specific analysis, I will analyze them separately.

2.2.1 Communist Legacies

The theory of communist legacy argues that the nature of the prior regime exercises constraining factors in the transition process of post-communist countries. The former communist countries are characterized by a diversity in the democratic development and party-system formation. The cause of such diversity lies in the different historical legacies of these countries (Kietschelt, 1999,pp.19; 2001, pp.299). It is a logic of cause and effect, where the past has influenced the nature of the communist regimes

considered as a 'passing phase in a longer trajectory of economic, institutional, and cultural development', and the nature of communist regime influenced the democratic institutions and party system formations (Kitschelt, 2001,pp.307).

Kitschelt et.al argue that the different types of communist regime – patrimonial, national-accommodative and authoritarian-bureaucratic – affected the choices of the political elite with regard to political institutions and party system competition, in the eve of the democratization process (1999, pp.383-384).

The countries that fall in the category of patrimonial communism, like Albania, had no dissidence during communism. In the transition period, a dissatisfied part of the communist elite, managed to organize itself and form opposition parties, before anticommunists could mobilize for a liberal-democratic order. The post-anti-communist parties produced weak programmatic crystallization. Their discourse was focused on anticommunism, thus framing the inter-party political competition not in terms of right-left dimension, but on the antagonism between apologists and opponents of the former communist regime (Kitschelt et.al, 1999, pp.388).

2.2.2 Path Dependency and Critical Junctures

For Mahoney 'path dependence occurs when a contingent historical event triggers a subsequent sequence that follows a relative deterministic pattern' (2000, pp.535). Through mechanisms of self-reinforcing or reactive sequence, this contingent historical event influences the outcomes of the future events, because it 'set into motion institutional patterns or event chains that have deterministic properties' (pp.508-508).

An intrinsic concept to the theory of path dependence is 'the critical juncture'. Pierson argues that 'junctions are "critical" because they place institutional arrangements on paths and trajectories, which are then very difficult to alter' (2004, pp.114), because institutions are enduring entities which are characterized by stability and difficult to change. Therefore critical junctures are essential to the understanding of path dependency.

Therefore it is not only important what happens but also the 'order of events' because the order 'affects how they happen'. It is important to mention that the critical juncture does not produce a simple chain of cause-effect. The agents have to choose among alternatives. Therefore, one cannot predict what the final outcome will be, based only on the initial condition (Mahoney, 2000, pp.511). However, that is true only on the initial phase, because 'once processes are set into motion and begin tracking a particular outcome, these processes tend to say in motion and continue to track this outcome' (Mahoney, 2000, pp.512).

The mechanisms of path-dependency are those of self-reinforcing and reactive sequences. Mechanisms of self-reinforcing are linked with 'increasing returns' where 'the probability of further steps along the same path increases with each move down that path' (Pierson, 2000, pp.252). It means that it is a process of institutional

reproduction. Once the change occurs it perpetuates itself by reinforcing the initial result (institution). However, there are also reactive sequences mechanisms, which occur against the established pattern, so that a new pattern could come to force (Mahoney, 2001, pp. 114-115). It could be said that the reactive sequences happen before self-reinforcing mechanisms take place.

2.3 Elite Cohesion and Ideological Crafting

Hanley et.al, contend that the elite cohesion and ideological crafting offer a better explanatory framework to explain the success of centre –right formations in Central and Eastern Europe. They view the elite cohesion and social political positioning as an explanatory framework for the success of the centre-right political parties in the post-communist countries of Central and Easter Europe. With elite cohesion is meant the 'ability of an elite group over time to reach and maintain consensus over key strategic and policy issues' (2008, pp.425). In their analysis of the Hungary, Czech Republic and Poland, they observed that the positioning of the successful centre-right party forming elites 'as credible second-rank challenger elites to the opposition elites who initially assumed power after 1989 when the unity of broad opposition groupings broke down in the early 1990s' (2008, pp.426, 427). The approach places more emphasize on agency rather than on structure, which is a different approach from that of macro-institutional and historical structural ones.

The application of the framework in the Albanian case requires the identification of elites, evaluation of their cohesion and identification of 'credible second - rank challenger elites'. The Democratic Party has experienced several internal crises since when it was created in late 1991. However, even though those who created splinter parties, were people of high profile with key positions within the Democratic Party, they never managed to be successful when competed in parliamentary and local elections. The electoral support of the Democratic Party was not hindered by its fragmentation. In the elections of 2005, all those who have left the party returned and participated in the parliamentary elections of 2005 as candidates of the Democratic Party. They competed in the majoritarian part and some of them didn't win, even though the party won the general elections. This is an indicator that neither the fragmentation of the party, nor the unification of the Democratic Party's elite influenced its electoral support. The emergence of second-rank challenging elites to the opposition elites that emerged in the aftermath of the collapse of communism did not happen in Albania. Therefore, we may conclude that the theory of Elite Cohesion does not provide a suitable framework in explaining the success of the Democratic Party.

Hanley et.al, consider that crafting of durable political ideologies is to be one of the two key factors in determining the success of centre-right formations in the postcommunist countries. 'Ideologies plays a crucial role in framing political action, giving cohesion and identity to political organizations and socializing incoming elites' (2008, pp.427). I will test such hypothesis in the Albanian case and see whether it has managed to be an 'ideological entrepreneur' able to 'develop an integrative ideological narrative that can provide a sustainable basis to develop a broad and durable political formation' (Hanley et.al, 2008, pp.427). In an analysis of the discourse of the Democratic Party, I have observed three phases: 1990-1993,1993-2004 and 2005- present. It is my hypothesis that when the discourse of the Democratic Party has been integrative the Party has performed successfully, while when the discourse has been exclusive, the party has performed unsuccessfully. Therefore, crafting an inclusive ideology proved to be a successful strategy. The Democratic Party managed to achieve the success only when it managed to provide a clear political platform for the electorate.

3. Proposed Methodology

This research is based on a vast literature that offers different explanatory frameworks that scholars have developed to analyze the emergence and development of political parties. The theoretical framework is based on the models of the structure and organization of political parties and on theories that explain their transformation. It also draws on the existing theories that explain the success or failures of the political parties in Central and Eastern Europe. The application of the models of party organization that have been developed in the Western European context to the parties in the new democracies has the risk to limit the degree of understanding of the processes of party development in CEE. However, these models are the only 'theoretical constructs available for the analysis of party organization, and it would be undesirable to dismiss them entirely' (van Biezen, 2003, pp.7). The study identified three existing approaches which tempt to explain the development of post-communist political parties. These approaches are: macro-institutional approach, historical structural approach and elite cohesion and ideological crafting approach. However, I consider as important to analyze the success of the centre-right party in Albania in the light of party typology as well.

Since the research is an attempt to analyze the success of centre-right in Albania, I use the conceptualization of Hanley et.al, for the definition of 'success' and 'centre-right'. The concept of ''party success' is defined in terms of: 'breadth' and 'durability'. With 'breadth [is meant] the ability to construct an inclusive electoral entity that encompasses a socially and ideologically broad range of voters and sub-groups' and by 'durability' [is meant] the ability of such an entity to remain united and cohesive and endure over a period of years' (2007, pp.6). The authors reject the elements of 'office-holding' and 'policy – or performance-based' measures of success, because they are determined by a complex array of economic, political, social and institutional factors (Hanley et.al, 2007, 6; Hanley et.al, 2008, pp.408). In conclusion, 'breadth and durability are conceptualized as the ability to garner a substantial proportion of the votes cast for all centre-right and right-wing party formations over a stained period of time' (Hanley et.al, 2007, pp. 9; Hanley et.al, 2008, pp.410).

The definition of the 'right' is based on: a) parties' self-identification; b) local understanding of 'rightness'; c) established patterns of coalition preference within national party systems; d) membership of transnational centre-right groupings such as the European People's Party or the European Democrats (Hanley et.al, 2007, pp.8; Hanley et.al, 2008, pp.410).

Similarly with Hanley et.al, (2007; 2008), in order to analyze breadth and durability I analyze:

1- The proportion of the vote for centre-right and right-wing parties taken by the largest centre-right or right-wing grouping;

2- the share of the centre –right and right-wing vote won by the larges centreright or right wing party divided by the number of right-wing parliamentary parties (the centre-right aggregation index);

3- the proportion of parliamentary seats won by the centre-right and right taken by the largest centre-right or right-wing grouping and;

4- the level of fractionalization of centre-right and right-wing forces in parliament as measured by the application of the Rae index to parties on the centre-right or right.

The research is based on primary and secondary sources. The primary sources consist mainly in the evaluation reports of the Organization for Security and Cooperation in Europe on elections in Albania; Central Election Commission of the Republic of Albania, International Foundation for Electoral Systems, Freedom House and EU Reports on enlargement. All of them were available online. Other primary resources are internal documents of the Democratic Party such as interim reports and meeting – minutes, as well as the official newspaper (Rilindja Demokratike) of the Democratic Party of Albania.

The secondary sources comprise the vast literature on the political parties in Western, Southeast and Central and Eastern Europe. Other secondary sources are some books and articles on the political parties, elections and transition in Albania. The books are in the Albanian language and were gathered during the period 07-21 July 2010, while I was in Tirana.

4. Conclusions

I started to write this piece of work with the aim of explaining the success of the Democratic Party of Albania. The research is conducted in a time where the Democratic Party of Albania is in the head of a coalition that for the first time in the Albania history is composed by right-wing and left-wing parties. The paper starts with the presumption that the Democratic Party is a successful one. The presumption is based on the fact that in 2009 the party won for the second consecutive time the elections and now is governing the country. It has won four out of seven parliamentary elections during the period 1991-2009. Another reason for such claim is that since its inception with exception to short terms, the Democratic Party has managed to build broad and durable party formations.

The factors that have contributed to the success of the Democratic Party are two. One is that the party has been able to craft an inclusive ideology. When the Democratic Party took an integrative ideological discourse it had a clear political platform, based on ideological lines, and this was reflected in a high number of supporters. On the other hand, the adoption of an excluding discourse which tends to envisage the opposition as enemy, was reflected in a low electoral support. I consider it as imperative to analyze the concept of 'anti communist rhetoric'. The 'anti communist rhetoric' of the Democratic of Albania has undergone three phases. The first phase was 1991-1993', where the anti-communist rhetoric was focused on concepts such as rights', 'individualism', 'free-markets', 'entrepreneurship' 'human and 'liberalization'. This phase coincides with the period when the DP came to power. The second phase begun after the radical reforms that Democratic Party took in economy. It encloses the period 1994-2004 when the DP had low electoral support. The rhetoric that characterizes this period was based more on excluding and demonizing the other or 'the enemy'. The third phase begun in 2005, which is the period that the DP came again in power. Its political platform is again on ideological lines. Therefore the Democratic Party achieved its success when it managed to provide a clear political platform for the electorate.

The second factor is linked with the structure of the party. The Democratic Party is a mass party and as such is disciplined and has penetrated in 'the deep fabric of the society' (Kajsiu, 20005, 143). There is a plethora of scholars who argue that the new democracies in the post communist countries are not experiencing the same path of party development that took place in the West. They claim that the mass party is passé and other models of party organization, such as the catch-all, professional-electoral or cartel could better suit to the political parties in Central and Eastern Europe (Lewis, 1996, pp. 10). Van Biezen, also argues that the parties in the new democracies experience weak partisan and strong electoral linkages, reduced relevance of party members and predominance of professional and party leaderships. The parties in new democracies have adopted the organizational style of the parties in the old democracies. According to her it is not the sheer newness of democracy what matters but the institutional context and the period in which parties develop (2003, pp.218-219). However, I do not agree such hypothesis. The emergence of mass parties, according to Duverger, is linked with the universal suffrage (1954, pp.65). The masse who were not allowed to participate in the political sphere, all of a sudden were allowed to do so. With the collapse of communism and the emergence of democracy, the people experienced the same phenomenon: universal suffrage. It is true that during the communist system, the citizens were entitled to vote, however the voting was commanded from the Party. Therefore the concept of voting was devoid of its meaning. With the collapse of communism, the process of voting took its meaning. The vote of the people had an effect on the political sphere. Hence, the need to integrate the people in the political organizations in a mass party fashion. However, it has resulted in creating patronage relations with its electorate (Kajsiu, 2008, pp.53-54). The loyalty of the members is rewarded with jobs in the public administration. All the reports of EU on Albania, as well as other organizations identify that there is a high level of political appointment in the public administration. A poll organized by the Ministry of Integration in 2007 showed that 28% of the staff was hired during the time that the Democratic Party came to power. Therefore patronage is another factor that explains the why Democratic Party of Albania is able to have a loyal electorate.

My argument is that the structure of the organization is important to the success or failure of a party. Therefore, being a mass party with an extensive membership that has penetrated deeply in society, has enabled the Democratic Party to implement its electoral strategies as well as ensure a loyal electorate. However, the relation of the party with its electorate consists in a system of distribution of incentives (jobs) in return for the loyalty. Therefore, the Democratic Party has ensured the loyalty of its supporters through patronage. What Democratic Party reaches through membership and patronage, the centre-right Parties of the CEE reach through elite cohesion.

In conclusion the Democratic Party of Albania owes its success to its ability to craft an inclusive ideology and to its structure as a mass party which enables it to create patronage and clientelistic links with its electorate.

References

Bartolini, S. and Mair, P. (2001) 'Challenges to Contemporary Political Parties', in Diamond, L. and Gunther, R. (eds.) *Political Parties and Democracy*, The Johns Hopkins University Press, Baltimore and London, 327-343.

Biberaj, E. (1998) The Rocky Road to Democracy, Westview Press: Boulder, Colorado.

Bugajski, J. (2002) *Political Parties of Eastern Europe – A Guide to Politics in the Post-Communist Era*, The Center for Strategic and International Studies, London.

Cili, H. (2002) Shumica që na mund (The majority that defeat us), Institute of National Studies, Tiranë.

Clark, T.D. (2002) Beyond Post-Communist Studies, M.E. Sharpe, New York.

Collier, R. B. and Coller, D. (1991) Shaping the Political Arena. Critical Junctures, the Labor Movement, and Regime Dynamics in Latin America', Princeton University Press, Princeton

Diamond, L. and Plattner, M. (2006) 'Introduction', in Diamond, L. and Plattner, M. (eds.) *Electoral System and Democracy*, The Johns Hopkins University Press, Baltimore.

Duverger, M. (1964) *Political Parties. Their Organization and Activity in the Modern State.* 3rd eds. Methuen & Co Ltd., London.

Fuga, A. (2008) Media, politika, shoqeria (Media, politics, society) 1990-2000, Botime Dudaj, Tirane.

Grzymala-Busse, A.M (2002) Redeeming the Communist Past. The Regeneration of Communist Parties in East Central Europe, Cambridge University Press, Cambridge.

Hanley, S. (2006) 'Getting the Right Right: Redefining the Centre – Right in Post-Communist Europe', pp. 9-27 in Szcerbiak, A. and Hanley, S. (eds.) *Centre-Right Parties in Post-Communist East-Central Europe*, Routledge, UK.

Hanley, S. (2006) 'Rise and Decline of the New Czech Right', in Szcerbiak, A. and Hanley, S. (eds.) *Centre-Right Parties in Post-Communist East-Central Europe*, Routledge, UK, 29-54.

Hanley, S., Szczerbiak, A., Haughton, T. and Fowler, B. (2007) 'Explaining the Success of Centre-Right Parties in Post-Communist East Central Europe: A Comparative Analysis', Sussex European Institute Working Paper No 94, [available at http://www.sussex.ac.uk/sei/1-4-10-1.html].

Hanley, S., Szczerbiak, A., Haughton, T. and Fowler, B. (2008) 'Sticking Together: Explaining Comparative Centre-Right Party Success in Post-Communist Central and Eastern Europe', *Party Politics*, 14, 407-434.

Kajsiu, B. (2004) 'Albanian Political Parties', pp.9-38 in Karasimenov, G. (eds.) *Political Parties and the Consolidation of Democracy in South Eastern Europe*, Gorex Press Publishing House, Sofia.

Kajsiu, B. (2005) 'Organizational Structure and Internal Party Democracy in Albania', in Krasimeonov, G. (eds.) *Organizational Structures and Internal Party Democracy in South Eastern Europe*, Gorex Press Publishing House, Sofia, 142-165.

Kajsiu, B. (2008) 'Ne rrenojat e nje utopia: Pse deshtoi Anti-Korrupsioni Shqiptar?, *Polis 6*, No.6 (Autumn, 2008), 42-57.

Kajsiu, B. (2010) 'Down with Politics! The Crisis of Representation in Post-Communist Albania', *East European Politics & Societies*, Vol.24 (2), 229-253.

Katz, R.S., Mair, P. (1995) 'Changing Models of Party Organizations and Party Democracy – The Emergence of the Cartel Party', *Party Politics*, 1 (1), 5 – 28.

Katz, R.S., Mair, P. (1996) 'Cadre, Catch-All, or Cartel? A rejoinder', *Party Politics*, 2 (4), 525 - 534.

Katz, R. S., Mair, P. (2002) 'The Ascendency of the Party in Public Office: Party Organizational Change in Twentieth – Century Democracies', in Gunther, R., and Monter, José Ramón and Linz, J. J. (eds.), *Political Parties – Old Concepts and New Challenges*, Oxford University Press, 113-135.

Kirchheimer, O. (1966) 'The transformation of the western European party system', in La Palombra, J. and Weiner, M. (eds.) *Political Parties and Political Development*, Princeton University Press, Princeton, 177 – 200.

Kitschelt, H., Mansfeldova, Z., Markowski, R. and Tóka, G. (1999) *Post-Communist Party Systems – Competition, Representation and Inter-Party Cooperation,* Cambridge University Press, Cambridge.

Kitschelt, H. (2001) 'Divergent Path of Postcommunist Democracies', pp.299 – 323 in Diamond, L. and Gunther, R. (eds.) *Political Parties and Democracy*, The Johns Hopkins University Press, Baltimore and London.

Koole, R. (1996) 'Cadre, Catch-All, or Cartel? A Comment on the Notion of the Cartel Party', *Party Politics* 2(4): 507 – 523.

Krasniqi, A. (2009) Zgjedhjet në Shqipëri (1991-2008). Historia, Proceset Zgjedhore, Partitë, Kandidatët, Fushatat Elektorale (Elections in Albania 1991-2008. History, Election Processes, Parties, Candidates, Electoral Campaigns), Tirane: Konica Kolor.

Krouwel, A. (2003) 'Otto Kirchheimer and the catch-all party', *West European Politics*, 26 (2), 23 – 40.

Krouwel, A. (2006) 'Party models', in Richard S Katz and William Crotty (eds) *Handbook of Party Politics*, Sage Publications, London, 249-269.

Lewis, P. G. (1996) 'Introduction and Theoretical Overview', in Lewis, P. G. (eds.) *Party Structure and Organization in East – Central Europe*, Edward Elgar, Cheltenham, 1-19.

Lewis, P. G. (2006) 'What Is the Right Way in East-Central Europe? Concluding Remarks', in Szcerbiak, A. and Hanley, S. (eds.) *Centre-Right Parties in Post-Communist East-Central Europe*, Routledge, UK, 133-148.

Lijphart, A. (1991) 'Constitutional Choices for New Democracies', Journal of Democracy, 2.

Lijphart, A. (1994) *Electoral Systems and Party Systems: A study of Twenty-Seven Democracies*, 1945-1990, Oxford University Press, New York.

Lipset, S. M. (2001) 'The Americanization of the European Left', in Diamond, L. and Gunther, R. (eds.) *Political Parties and Democracy*, The Johns Hopkins University Press, Baltimore and London, 52-66.

Mahoney, J. and Snyder, R. (1999) 'Rethinking Agency and Structure in the Study of Regime Change', *Studies in Comparative International Development*, Summer 1999, 3-32.

Mahoney, J. (2001) Path-Dependent Explanations of Regime Change: Central America in Comparative Perspective', *Studies in Comparative International Development*, Spring 2001, Vol.36, No.1, 114-141.

Mahoney, J. (2000)"Path Dependence in Historical Sociology," *Theory and Society*, 29, 507-548.

Nohlen, D. (1998) "Electoral Systems in Eastern Europe: Genesis, Critique, Reform," in *Electoral Systems, Political Stability and Viable Government in Europe's New Democracies –Achievements, Failures and Unresolved Issues*, eds. Heinz, W. and Klein, H. (Strasbourg: Proceedings of a Seminar at the Council of Europe, 22-25 November 1998).

Norris, P. (2004) *Electoral Engineering: Voting Rules and Political Behaviour*, Cambridge University Press.

Panebianco, A. (1988) Political Parties – Organization and Power, Cambridge University Press.

Pierson, P. (2000) 'Increasing Returns, Path Dependence, and the Study of Politics, *American Political Science Review* 94, 251-267.

Pierson, P. (2004) Politics in Time, Princeton University Press.

Rilindja Demokratike (1991) 'E duam Shqiperine si Gjithe Evropa (We want Albania to be like the rest of Europe', Rilindja Demokratike (Democratic Renaissance), No.2, 4.

Sartori, G. (1968) 'Political Development and Political Engineering', in J.D. Montgomer and A.O.Hirschman (eds) *Public Policy*, 17, 261-298.

Sartori, G. (1976) *Parties and Party Systems: A Framework for Analysis*, Cambridge University Press.

Sartori, G. (1986) 'The Influence of Electoral Systems: Faulty Lines or Faulty Method? In Grofman, B. and Lijphart, A. (eds.) *Electoral Laws and Their Political Consequences*. New York: Agathon Press.

Sartori, G. (2001) 'The Party Effects of Electoral Systems', in Diamond, L. and Gunther, R. (eds.) *Political Parties and Democracy*, The Johns Hopkins University Press, Baltimore and London, 90-105.

Schmitter, C. P. (2001) 'Parties Are Not What They Once Were', in Diamond, L. and Gunther, R. (eds.) *Political Parties and Democracy*, The Johns Hopkins University Press, Baltimore and London, pp.67-89.

Szcerbiak, A. and Hanley, S. (2006) 'Introduction: Understanding the Politics of the Right in Contemporary East-Central Europe', in Szcerbiak, A. and Hanley, S. (eds.) *Centre-Right Parties in Post-Communist East-Central Europe*, Routlege, UK, 1-8.

Taras, R. (2007) 'Executive Leadership', in White, S., Batt, J. and Lewis, P. G. (eds) *Developments in Central and East European Politics 4*, Palgrave Macmillan, New York, 127-144.

Van Biezen, I. (2003) *Political Parties in New Democracies – Party organization in Southern and East-Central Europe*, Palgrave MacMillan, Hampshire.

Vickers, M. and Pettifer, J. (1997) *Albania. From Anarchy to a Balkan Identity*. New York University Press, New York.

Corporate Social Responsibility: A Comparison Between Large and Medium Companies in UK

Maryam Sadeghi

PhD Student (Management School), University of Sheffield,

m.sadeghi@sheffield.ac.uk

Abstract. The current paper attempts to find out how various companies according to their size adopt corporate social responsibility practices. This research discusses how companies in different business sectors and size, participate in social practice activities. In addition it explores how companies define social practices in their business context. Furthermore, this research is on the path of finding how particularly business environment affect companies understanding of corporate social responsibility and their social activities.

Keywords: Corporate Social Responsibility, Medium-Sized Enterprise, Critical Discourse Analysis

1 Introduction

In recent years there has been great awareness and attention towards corporate social responsibility in the world. Corporate social responsibility includes a wide range of activities in various aspects of social, environmental issues. This reflects a significant role of companies in adopting appropriate CSR practices. Since one main responsibility of companies is making profit; balancing between business profit and social/ environmental responsibilities of companies, has been controversial issue in

corporate social responsibility. Always there has been a notion that there are other purposes behind social practices in companies (eg. Porter, Kramer, 2006; Roberts, 2003; Margolis and Walsh, 2002). In similar vein there have been various ideas about how companies are responsible towards society and environment. One controversial issue in this regard has been the size of companies and different characteristics of companies according to their size, for example there have been many studies that large companies have (more) reasons for adopting CSR practices, or smaller enterprise have different organizational characteristic that hinder CSR activities (Hitchens, et. al. 2005, Lepoutre, Heene, 2006). In contrast, some scholars believe that small and medium enterprise are interested in CSR and involved in related practices (Castka, et al. 2004, BITC, 2002, Lepourte, Heed, 2006). As there have been various ideas about how companies perform CSR activities and adopt social practice accordingly (Mingoja, 2008, Perrini 2006, Lepoutre, Heene, 2006, Observatory of European SME, 2002), it seems there is still a gap in CSR literature in order to find a deeper understanding of CSR practice natures inside the companies. Furthermore, this understanding can help to improve and develop CSR practices among companies, especially SMEs that have significant impact on economy, environment, and social aspects in the UK (Udayasankar, 2007, Castka et al. 2004, Simpson, et. al. 2004, Hitchens, et al. 2005, Lepoutre, Heene, 2006, Fassin, 2008, Simpson, et al, 2004, Hitchens, et al. 2005). Therefore for deep understanding of motivations behind Corporate Social responsibility and for developing to further level, this research aims to study how companies adopt CSR practices. This research is using qualitative methodology to add more understanding on CSR practice. The research samples are chosen from medium sized and large companies while there is justification how the samples for medium sized companies represent SMEs. In fact, this paper in an attempt to find out an acceptable basis for comparing Medium sized companies and large companies in CSR practices through which companies communicate and share their information with broad society. At this stage, samples are gathered from various sections of industries for both medium sized and large companies. Moreover, industries categorized according to their impacts on environments and climate change. Additionally, there was a careful consideration towards various aspects of corporate social responsibility inside companies. This paper analyses how organizations apply discourse in their social practice disclosure. According to Foucaultian theory it can be expected that organizations attempt to provide a meaningful context in order to convey their message to their customers and society (see Livesey, 2001, 2002a, 2002b). Additionally, these messages are based on interconnection between language and power in the context of organization in relation with broader society (Foucault, 1984, Livesey, 2001). Similarly, while language is a mean of organization power, it also indicates how organizations as social actors understand social phenomena (Halliday, 1973) or how organizations apply discourse in an understandable pattern by society (see van Leeuwen 2005, Chaiyasuk, 2007). In addition, for

better understanding of organizational practice, it is helpful to notice how discourses are adopted for organizational representation, persuasion and for the purpose of organizations; including social identity and social acceptance (see Van Dijk, 1990, Wodak, 2001). Specifically, the author examines responsible business practices of organizations to describe organizational rational for producing or increasing social and particularly stakeholder communications and reporting. Drawing on discursive domain of business accountancy and organizational value, this paper attempt to show how organizations apply discourses to reconstruct the interaction between the organization and society and the way organizations manage situations in social context (see Campbell and Beck, 2004). Alternatively, this paper explores meaningful concept that organizations try to make by discourses that is helpful to lead us to understandable business practices of organizations. This study, with using Foucauldian theory and Fairclough framework based on combination of Foucault theory and interactions within the context, attempts to interpret organizational social disclosure through Communication means and social reports. This systematic analysis will be in two levels of text, content and context. Accepting that this research is specific in context, there is a comparison between Medium Sized organizations and large organization in terms of using discourse for their social practice disclosure. There is little research in SMEs social disclosure, therefore this research is important in developing the researchers understanding of discourse using by various social actors in organizational context. Additionally, there is not enough research on investigating communication means that organization use to communicate its intention with society (Chaiyasuk, 2007, Caldas-Coulthard, C. R. 2005). This research focuses on textual representation and semiotic constructions to present a new insight to social disclosure defined by organizations including intentions and attitudes (see Wilson 1993, Morris, 1994, Krippendorff, 2004). Moreover, it attempts to present that organizations are interested in to represent what aspect of reality (Van Leeuwen, 2005) and how organizations apply specific textual feature (Fairclough, 1992, Chaiyasuk, 2007). Similarly, this research seeks social disclosure of organizations in order to interpret how organization deals with the issue or giving a general view of organization policies and procedures (see Campbell and Beck, 2004).

2 Methodology

The current research has two main aims, first to assess corporate social responsibility practices in companies and second to assess how these companies communicate their message behind corporate social responsibility through discourses. In order to understand the CSR practice I use content analysis. Content analysis can assist to find out 'intention', 'attention' and

analysis. Content analysis can assist to find out 'intention', 'attention' and 'attitude' towards specific issue in a systematic approach (Morris, 1994). Majority of content analysis are in quantitative approach and there is lack of enough studies that apply qualitative approach to content analysis (Tregidga et. al, 2007). Qualitative research provides a deep understanding of the analyzed data. Qualitative approach is mostly data based rather than being theory driven (see MacFarlane and Brun, 2011). The samples are selected purposive because the aim of qualitative approach is not generalizability but it explores the issues to understand phenomena (Morris, 1994, Krippendorff, 1980, 2004). Samples are gathered from various sections of industries for both medium sized and large companies. Moreover, industries categorized according to their impacts on environments and climate change. In a pilot study the annual reports of one specific industry were selected then the themes with inductive approach were defined. The next step will be critical discourse analysis of the selected samples. In this section, environmental/ social message of companies will be interpreted (See Livesey, 2001, 2002a, 2002b)

References

Campbell, D. and A. Cornelia Beck 2004. "Answering allegations: the use of the corporate website for restorative ethical and social disclosure." Business Ethics: A European Review 13(2-3): 100-116.

Caldas-Coulthard, C. R. (2005). Personal web pages and the semiotic construction of academic identities. The writer's craft, the culture's technology. C. R. Caldas-Coulthard and M. Toolan. Amsterdam, Rodopi: 23-46.

Castka, P., M. A. Balzarova, et al. (2004). "How can SMEs effectively implement the CSR agenda? A UK case study perspective." Corporate Social Responsibility and Environmental Management 11(3): 140-149.

Chaiyasuk, 2007, A critical description of the persuasive discourse of British and Thai university web pages. Thesis submitted for Doctoral Philosophy degree, University of Birmingham.

Fairclough, N. 1992. "Discourse and Text: Linguistic and Intertextual Analysis within Discourse Analysis." Discourse & Society 3(2): 193-217.

Fassin, Y. (2008). "SMEs and the fallacy of formalising CSR." Business Ethics: A European Review 17(4): 364-378.

Fassin, Y. and A. Van Rossem (2008). "Corporate governance in the CSR and ethics debate: sensemaking of experts and CEOs about social issues in management." Academy of Management

Foucault, M., 1972, The archaeology of knowledge and the discourse of language, New York, Pantheon Books.

Foucault, M., 1977, discipline and punish, the birth of the prison, New York, Random House.

Foucault, M., 1980, The history of sexuality, New York, Pantheon Books.

Foucault, M. (1980). Power/knowledge: Selected interviews and other writings 1972-1977, (C. Gordon, Ed.). New York: Pantheon.

Foucault, M. (1984). The Foucault reader (P. Rabinow, Ed.). New York: Pantheon.

Halliday, M. A. K. (1973). Explorations in the Functions of Language, London, Edward Arnold.

Hitchens, D., S. Thankappan, et al. (2005). "Environmental performance, competitiveness and management of small businesses in Europe." Tijdschrift voor economische en sociale geografie 96(5): 541-557.

Krippendorff, 2004, Content analysis: an introduction to its methodology, Sage Publication.

Lepoutre, J. and A. Heene (2006). "Investigating the impact of firm size on small business social responsibility: a critical review." Journal of Business Ethics 67(3): 257-273.

Livesey, S. 2001. "Eco-Identity as Discursive Struggle: Royal Dutch/Shell, Brent Spar, and Nigeria." Journal of Business Communication 38(1): 58-91.

Livesey, S. 2002a. "The Discourse of the Middle Ground." Management Communication Quarterly 15(3): 313-349.

Livesey, S. 2002b. "Global Warming Wars: Rhetorical and Discourse Analytic Approaches to ExxonMobil's Corporate Public Discourse." Journal of Business Communication 39(1): 117-146.

Livesey, S. and K. Kearins (2002c). "Transparent and Caring Corporations?" Organization & Environment 15(3): 233-258.

Tregidga, H. et. al. (2007). The Role of Discourse in Bridging the Text and Context of Corporate Social and Environmental Reporting

MacFarlane, A and Brún, M. (2011), Using a Theory-Driven Conceptual Framework in Qualitative Health Research

Margolis, J. D. and J. P. Walsh (2003). "Misery loves companies: Rethinking social initiatives by business." Administrative Science Quarterly McWilliams

A., D. S. Siegel, et al. (2006). "Corporate Social Responsibility: Strategic Implications*." Journal of Management Studies

Morris, R., 1994. "Computerized Content Analysis in Management Research: A Demonstration of Advantages & Limitations." Journal of Management 20(4): 903-931.

Porter, M. E. and M. R. Kramer (2006). "Strategy and society." Harvard Business Review

Simpson, M., N. Taylor, et al. (2004). "Environmental responsibility in SMEs: does it deliver competitive advantage?" Business Strategy and the Environment 13(3): 156-171.

Udayasankar, K. (2008). "Corporate social responsibility and firm size." Journal of Business Ethics 83(2): 167-175.

Van Dijk, 1990, "Social Cognition and Discourse" Handbook of Language and Social Psychology Edited by H. Giles and W.P. Robinson, John Wiley & Sons Ltd

Van Leeuwen, T., 2005, "Introducing Social Semiotics", Taylor & Francis e-Library.

Wilson, A. 1993, Automatic content analysis of spoken discourse: a report on work in progress.

Wodak, R. 2001, "What CDA Is About - a Summary of Its History, Important Concepts and Its Developments." In Methods of Critical Discourse Analysis, ed. R. Wodak and M. Meyer, 1-13. London & Thousand Oaks & New Delhi: Sage Publication.

Neo-Gramscianism and European Integration

Mihail Caradaică1

Abstract. This paper belongs to the area of critical studies in European Integration and tries to demonstrate that neo-gramscianism is an integration theory which came as a necessity to complete the puzzle in this field. The main critique of neo-functionalism, intergovernmentalism and middle-range theories is that they are unable to understand the real nature of power in the European Union, and by this I mean that they cannot conceptualise the power relations which are part of capitalist market structures. In other words, those theories fail to understand the structural power that establishes the direction of the European integration.

My purpose here is to show the limits of the classical² and middle-range theories and that the neo-Gramscian theory covers the gaps of the traditional integration ones, providing a coherent vision on the emergence and the evolution of the European Union, by using the social forces agency in the process of integration and superstructural dimension of European Single Market. I choose this critical theory because it focuses on the process of change, adopting the historical-structures perspective, which regards all of the structures that define social and political reality (ex. human nature) as products of history and subjects of change.

Keywords: European Integration, neo-Gramscianism, Social forces, Hegemony

1. Introduction

In this article I will try to outline a/the neo-Gramscian theory of European integration starting with a critique on the classical theoretical debate. In all the issues where the traditional theories fail to explain reality, I will approach the neo-Gramscian concepts and formulate a coherent understanding of the European integration process. The main argument that a new theory is needed, lays in the variation of theoretical approaches that characterised the stages of European integration since the emergence of the European Coal and Steal Community.

"When the integration process was going well, as during the 1950s and early 1960s, neo-functionalists and other theorists sought to explain the process whereby European integration proceeded from modest sectoral beginnings to something broader and more ambitious" (Pollack 2010, p. 17). But between the 1960s and the 1980s, when due the oil crisis the European integration lost its strength, "intergovernmentalists and others sought to explain why integration process had not

¹Mihail Caradaică is a PhD candidate at the National School of Political Studies and Public Administration. E-mail address: mihai.caradaica@gmail.com.

²The classical theories are neo-functionalism, intergovernmentalism and liberal intergovernmentalism.

proceeded as smoothly as its founders had hoped" (Pollack 2010, p. 17). And if we look further, nowadays middle-range theories already give up understanding the whole European integration process, being focused on specific aspects like policy-making.

I assume that the neo-Gramscian framework of analysis is capable of understanding the process behind the political decisions that represent important steps in the European integration process. I am also aware of the critics that combat neo-Gramscianism with its own weapons, saying that the author's "thought needs to be historicised and understood against the background of his own time" (Bieler 2000, p. 9). But Gramsci himself believes that studying the history and appropriating ideas is a way of understanding the present.

2. Literature review

2.1. The limits of classical debate

For a more detailed view on the limits of the classical and middle-range theories, I will present a number of approaches. The first critique is presented by Bastiaan van Apeldoorn, Henk Overbeek or Magnus Ryner, claiming that "these mainstream³ theories fail to account for the *structural power* that determines the particular trajectory of European integration" (Apeldoorn, Overbeek and Ryner 2003, p. 17). Another critical point of view belongs to Andreas Bieler, who assume that for all of those theories the "integration starts when it is realised that certain economic problems yield higher welfare gains, if they are dealt with at the supranational level" (Bieler 2000, p. 4). Both visions will be more detailed in the next sections.

2.1.1 Neo-functionalism

The neo-functionalism theory emerges in 1958 with Ernst Haas book called The Uniting of Europe: Political, Social and Economic Forces, and it is based on several assumptions. The first one is the concept of spillover which is the core of this theory. Haas theory "was based on the assumption that cooperation in one policy area would create pressure in a neighbouring policy area, placing it on the political agenda, and ultimately leading to the future integration" (Jensen 2010, p. 72). In other words, it means that cooperation in one field would create the necessity of cooperation in other fields. The second one refers to the role of societal groups in the integration process. "Haas argued that interest groups and political parties would be key actors in driving integration forward" (Jensen 2010, p. 73). It means that those groups will see the integration process as a way to resolve different problems they face. And the third one is about development of supranational loyalties by the European elite. "Thus neofunctionalists predict that the European integration process would lead to the establishment of elite groups loval to the supranational institutions and holding pan-European norms and ideas" (Jensen 2010, p. 77). This elite will afterwards try to influence and convince national elites of the advantages generating by the integration process.

Also I cannot avoid here the importance of supranational institutions which will escape from their creator power and will give a different direction to the

³ For Apeldoorn, Overbeek and Ryner mainstream theories symbolise the same thing as classical theories.

European integration process. "Concerned with increasing their own powers, employees of regional institutions become agents of further integration by influencing the participation of participating elites (both private and public), and therefore governments' (national) interest" (Niemann and Schmitter 2009, p. 48). From the mid-1960s this theory was increasingly criticised because of the empirical events taking place in European Union politics, like the crisis of the Empty Chair generated by the French President Charles de Gaulle or during the oil crises from the 70s and 80s when the integration stopped. Thus, "in the mid-1970s, Haas declared the theory to be obsolete"(Niemann and Schmitter 2009, p. 45).

The first critique regarding neo-functionalism is that it cannot focus on the internal and external power of interest groups, and because of this it cannot understand why some groups are more powerful than others in the process of establishing the European Agenda (Apeldoorn, Overbeek and Ryner 2003, p. 22). Even if it recognises their existence and importance, this theory is not able to offer many details about the nature of these groups. Apeldoorn continued the argument and said that "nor did neo-functionalism pay much attention to the question of power within and between groups, of why some groups are more powerful than others and may thus be more successful in setting the agenda of European integration" (Apeldoorn 2002, p. 36). It means that there is no explanation for which groups should succeed coalitions, mobilise interests or gain access to the decision making process and influence policies. Also, in the neo-functionalist theory the dynamics of interest groups were seen as coming into existence as a functional response to the requirements of the supranational system⁴"(Apeldoorn 2002, p. 35).

Another problem of the supranationalism in general is that many aspects remain under-theorized. For example, "there is no explanation of where transnational interests come from and why they would be so powerful" (Apeldoorn, Overbeek and Ryner 2003, p. 23), and this is possible because no explanation of how social forces shape the socio-economic environment is available. Also, transnational society is understood as a concept, and there is no theory to explain it. Moreover, "the role of transnational business is seen as a response not only to the international environment but also to the institutional changes (i.e., the SEA) that were already taking place, leading to a new reality that was perceived as 'irreversible'" (Apeldoorn 2002, p. 37). In this logic, Big business played a role in the insertion of the Single Market on European Agenda, but this role is understood by the neo-functionalists like a support to political initiatives that already exists.

Bieler completes the critique and finds two big problems. The first one is based on an ahistoric understanding of the human being which is supposed to be rational and formed by utility-maximising individuals. According to this, "the notion of spill-over implies an inevitable, teleological process of further integration along an objective economic rationality" (Bieler 2000, p. 4). Secondly, neo-functionalism "explains European integration through an emphasis on the internal dynamics of

⁴ This is obvious a wrong generalisation about the process of European integration because it is impossible to explain the whole process in a same direction. Neo-gramscianism focus to explain historical situation because certain events could happened as a structural changes in European economy or as a pressure from super-structural dimension.

926

European politics" (Bieler 2000, p. 4). But, by doing this, it totally neglects the wider structure in which the European Union is situated (ex. globalisation).

2.1.2 Intergovernmentalism

Intergovernmentalism, represented by Stanley Hoffman, is a European integration theory with the roots in international relations realism. The intergovernmentalism theory is characterised by the state-centrism and privileges the role of the state in the European integration. Hoffman theory has its roots in international relations realism, and claims that there is no global authority, state relations being dominated by anarchy. He "accept that international institutions of all kinds are established to reduce the level of anarchy within the state system, and see the European Union as just another of these institutions" (Cini 2010, p. 89). According to this type of thinking, there are costs and benefits of cooperation and a state will cooperate if the benefits are higher than costs. In these respect, intergovermnetalists prefer to talk about European cooperation rather than European integration. Furthermore, "rather than assuming that these institutions are capable of playing an independent or autonomous role within the European integration process, intergovernmentalists tend to stress that the so-called supranational actors, the Commission in particular, are little more than the servants of the member states" (Cini 2010, p. 90). It means that there is no institutional path dependence and the integration process is reversal.

The first critique is about the absence of society from the analysis. Thus, state interests are determined by the balance of power from the international anarchic system, while inside it is still a *black box*. The critique here is that "These national policy preferences are, however, not explained, but are taken as given" (Apeldoorn, Overbeek and Ryner 2003, p. 23). "Not only are transnational actors ignored in the state-centric ontology of intergovernmentalism, but social forces are excluded altogether from the analysis. The state is the central actor, and the state is seen as separate from society"(Apeldoorn 2002, p. 37). In this regard, the Single European Act, for example, is explained just as an intergovernmental bargain settled by the national preferences which are not explained.

For Bieler, the classical intergovernmentalism is easier to be criticised. Even if it pays attention to the international structure, it is focused on the states as the only actors which produce change. Thus, "it cannot account for structural changes such as globalisation, which go beyond the state structure" (Bieler 2000, p. 4).

2.1.3 Liberal intergovernmentalism

The liberal intergovernmentalism of Andrew Moravcsik opens the state black box, establishing a theory of the national preferences formation. At once, Moravcsik remains an intergovernamentalist regarding inter-state relations. However, "drawing on liberal theories of national preference formation, and applying a domestic politics approach, Moravcsik shows how state goals can be shaped by domestic pressure and interactions which in turn are often conditioned by the constrains and opportunities that derive from economic interdependence" (Cini 2010, p. 97). To be more clear, with liberal theories Moravcsik explains the formation of national preference by the pressure of domestic social actors. Then, we have the configuration of the state preferences at national level, and after, international negotiations seen through intergovernmental theory. In this last process the most important things are the intensity of national preferences, the alternative coalitions and available issue linkages.

Even if this theory comes with an innovative model of explaining the European integration, it has a lot of unexplained gaps. Moravcsik analysis does not go beyond the weaknesses and power of what he calls *production groups*, although it is obvious that they are favoured in front of consumers, tax payers or third world producers. In other words, "there is no interest in the structural inequalities that are constitutive of the balance of social forces and how these forces change over time" (Apeldoorn, Overbeek and Ryner 2003, p. 25). "Liberal intergovernmentalism is hampered, then, by a rather poor notion of state-society relations, very much modelled on that of neo-classical economics) (Apeldoorn 2002, p. 40). Society, in this case, is nothing more than an aggregation of individuals and interest groups which are acting in the logic of cost-benefit analysis. Also, this view is obviously neglecting the ideological motivations of the actors.

Furthermore, the theory is bounded by state-society relations, without considering any historical approach about the emerging social relations: "There is no conception here of how historically constructed *social relations* embed this actors and shape their identity and interests. Society is thus emptied of most of its historically produced content. If we want to understand the social purpose underpinning European integration, such an atomistic conception of society is, I would contend, not very helpful" (Apeldoorn 2002, p. 40). Another problem with Moravcsik theory is that "social forces are contained within the boundaries of national states" (Apeldoorn, Overbeek and Ryner 2003, p. 25-26), which is almost an impossible issue in a European society that becomes more competitive and integrated in global economy.

On the other hand, Bieler, the liberal intergovernmentalism, which tries to take into account interest groups, is instead focusing on each country's domestic realm. It suffers grave deficiencies by not providing an "insight into how the independent role of ideas is to be investigated or how transnational actors can be accounted for" (Bieler 2000, p. 6). The actions of all interest groups are limited to the national societies. And to conclude, regarding the liberal intergovernmentalist framework, "a proper understanding of the changing social purposes underlying European order is moreover constrained by the fact that this theory remains very state centric, to the extent that states remain the only important actors within world and European politics" (Apeldoorn 2002, p. 40).

2.1.4 Middle-range theories

The new approaches of European integration, middle-range theories, are by definition excluded from my analysis. "As the name suggest, middle-range theories do not have totalizing ambitions; they seek to explain aspects of a phenomenon rather than its whole" (Rosamond 2010, p. 108). My aim in this article is not to describe aspects of European integration, but the entire process. According to this, middle-range theories are no longer relevant for this paper. Also, Bieler observes that the middle-range theories of European integration, besides the fact that they do not aim to explain the entire integration process, are also focusing on a variety of actors that could influence the decision making process at national and supranational level. By doing this, those authors make a dangerous distinction between European Integration

928

and Comparative Politics. "The two areas are closely connected and should not be separated analytically" (Bieler 2000, p. 8).

2.2 Neo-Gramscianism as a theory of European integration

The limits of the classical European integration theories have been located by Marxist authors since the European Community started to function. I would like to point out these three authors: Ernest Mandel, Nicos Poulantzas and Peter Cocks. The first two authors focus their research on explaining some particular aspects of European integration focusing on the transformations of capitalist system.

Mandel is seeing the emergence of Common Market as a product of capitalist concentration on an international level which tries to maintain state sovereignty: "an attempt by capitalism to reconcile the level of development of the productive forces and the degree of monopolistic concentration with the survival of national state" (Mandel 1967, p. 27). Thus, he identified three types of capitalist concentration specific to that time in the Common Market. The first one is the fusion of the national enterprises to compete with other enterprises of Common or international market. Secondly, there is the fusion or absorption (in most cases) of Common Market national companies by other American companies. And thirdly, "the fusion of national companies of various Common Market countries into new units in which national capital is no longer predominant, but in which capital is now more or less equally dispersed over two, three or more Common Market countries" (Mandel 1967, p. 28). All those transformations of economical structure caused a series of superstructural changes materialized in European political institutions. In other words, "the appearance of large amalgamated banking and industrial units which are not mainly the property of any national capitalist class, represent the material infra-structure for the emergence of supra-national state-power organs in the Common Market"(Mandel 1967, p. 31).

As Mandel state, this situation proofs "the old Marxist *dictum* that in our epoch the productive forces have obviously outgrown both the boundaries of private property and of the national state" (Mandel 1967, p. 29). Furthermore, he identified a key moment that could happen at any time, and that could deeply affect European Community: a recession. "When a general recession breaks out in all the six countries (and this seems to us inevitable), the "moment of truth" for the common market will arrive" (Mandel 1967, p. 32). Thus, there are practically two exits from this situation: returning to economic nationalism and protectionism or by enforcing supra-national institution to manage the situation (which means to deepen the integration process). In the second case, says Mandel, European Community could face a single currency or a common taxation system. Considering this, the choice between those directions will be influence by the relative strength of the interested bourgeois forces that could be nationals or supra-nationals.

In my opinion, this point of view is limited because it doesn't take into consideration the possible influence of super-structure⁵ in the future development of European Community/European Union. But the general directions of evolution given by this author were legitimated by the future development of EU. During the oil crisis

⁵ This aspect is going to be solved by the neo-Gramscian view, which is focus on both structural and super-structural dimension of European integration.

of 1973 and 1979 member countries turn back to their national economy trying to protect it by individual agreements with other economical agents. Thus, European integration was not a priority for almost 20 years. On the other hand, during the financial crisis started in 2008 and due the transformation of economic structure⁶, member states did not react like 30 years ago and this time supra-national common policies (for example European Fiscal Pact) replaced individual strategies.

Nicos Poulantzas is more focused on structural arrangements and analyses the American capital hegemony on Europe. He affirming, in an article called "Internationalization of capitalist relations and the nation state" and wrote in 1974, that we are facing a quasi-disappearance "of the powers of the national state of the imperialist metropoles, be it under the domination of the American superstate or under the domination of large American Capital or of *international* capital liberated from the shackles of the state" (Poulantzas 2008, p. 221). But even if European capital would have been developed at that time, and it have had its own base for accumulation, it would have, inevitably, become more depended on American capital because of the US investments. Hereby, "The ECC and United Europe are considered the manifestation of an increased domination of American capital" (Poulantzas 2008, p. 222). But the question that Poulantzas didn't answer is why the nationality of the capital is so important and why the capital should have a nationality? My point of view here is that the only rule of capital is the pursuit of profit.

The most important of them, in my opinion, is Peter Cocks due to his article "Toward a Marxist Theory of European Integration" written in 1980. In this article, he tried to discover if and why integration occurs using a historical analysis of the capitalist development and also tried to explain, from a Marxist point of view, the whole integration process.

His analysis begins with a critique of the orthodox integration literature which is called fundamentally ahistorical, because of its incapacity for providing an adequate account of the roots of nowadays European integration. "Consciously or not, they avoided the question of whether integration is qualitatively different in different socio-economic formation, why it emerges at some historical periods and not others, and what is the connection between different levels of integration in distinct social systems" (Cocks 1980, p. 2). Following these basic questions, the duty of Marxist scholars is to investigate why specific patterns of integration arose in specific European eras. Cocks tries to do this and state how those social, political and economic integration dynamics are shaped by the capitalist system. The configuration of these three phenomena, in time, "depends at minimum on the specific phase of capitalist development, the technological and administrative state of knowledge, the level of political consciousness of the masses, and the perception and activities of the dominant political and economic classes" (Cocks 1980, p. 35).

Since its beginning, capitalism presented distinct practices like: working in a wage system, private ownership of the means of production or the fragmentation of life. But more important, it created an ideology to justify its activities and to maintain its power structure. "The bourgeois mind thus to endorses individualism, the holding of private property, market relations, acquisitiveness, competition and profit, which

⁶ During the years 1970 and 1980, is ws a transition period from Fordism to post-Fordism type of capitalism.

are incorporated as values in everything from law, to education, to religion, to literature" (Cocks 1980, p. 35). Gramsci calls this "ideological hegemony" and it represents the next step in my research.

The neo-Gramscian theory continues the Marxist tradition in European integration studies and focuses on two central concepts: super-structural⁷ dimension and social forces agency. It tries to capture a so-called real picture of the integration by showing the forces which shape the European decision-making process (market forces) and the nature of ideological hegemony.

Adam Morton, a neo-Gramscian author, wrote that "hegemony within the realm of civil society⁸ is then grasped when the citizenry come to believe that authority over their lives emanates from the self" (Morton 2007, p. 93). In other words, hegemony does not represent, like in the neorealist view, a hegemonic state that controls and dominates other states, but "type of rule which predominantly relies on consent rather than on coercion" (Bieler 2000, p. 14). Furthermore, the dialectical relation between economic structure and ideological super-structure produces, as Gramsci states, the historical bloc. In classical Gramscianism, the historical bloc is the alliance between the working class and the bourgeoisie, accomplished through the cultural hegemony. But a historical bloc is more than a mere alliance between social forces. It is "the term applied to the particular configuration of social classes and ideology that gives content to a historical state" (Robert Cox (1987). "Production, Power And World Order: Social Forces in the Making of History", New York: Columbia University Press, p. 409 in Bieler 2000, p. 14). In nowadays European Union, the ideological hegemony is represented by neo-liberalism. But as I already mentioned, this specificity of super-structure should be based on a specific arrangement inside the economic structure, which is shaped by social forces. To conclude, "a hegemonic order is based on a historical bloc that does not necessarily coincide with the boundaries of a state, but may be established at a transnational level" (Bieler 2000, p. 14).

I should add that between structure and super-structure there is a reciprocal relationship, that for Gramsci is nothing more than a dialectical process. "Superstructures of ideology and political organisation shape the development of both aspects of production [that is, the social relations and the physical means of production] and are shaped by them" (Robert Cox (1983). "Gramsci, Hegemony and International Relations: An Essay on Method". *Millennium: Journal of International Studies* pp.162–175. p. 168 in Bieler 2000, p. 14)

The social relations, or force relations as Gramsci called them, are in fact relations between different social groups. Thus, the relations of force operate on three

⁷ In Marxist theory, structure engulfs all the social relations of production, or the economic base. Super-structure symbolizes the effect of economic arrangement of a society in ideology – ideological super-structure.

⁸ Gramsci divides the society in two parts: civil society and political society. "Political society includes the public sphere of government, administration, and law and order, as well as security. Civil society includes those elements normally considered private, such as free enterprise, political parties, churches, and trade unions" (Gill 2003, p. 65).

interconnected levels: structural⁹, political¹⁰ and strategic¹¹. Methodologically, the object of the analysis is the *historical situation*, and the method of analysis is the observation of force relations (Gill 2003, p. 51). In other words, the dynamic of force relations produces certain historical events. For example, it is difficult to elaborate a general theory which could explain in the same time the emergence of the European Community¹² and European Union¹³.

For a better understanding of social forces, the nature of capitalism should be presented. As it is obvious, the capitalist mode of production is based on private enterprise and wage labour, which define the opposition between property owning and free labour. "Labour and capital are, consequently, two collective actors opposing each other, engendered by the production process as social forces" (Bieler 2000, p. 10). And because the production process is no longer organised at national level, capitalist accumulation exceeds national boundaries and social forces become transnational. But all this time, capitalism like current mode of production, changed its nature and according with this we have two phases in the European socioeconomic order. The first one is Europe after Second World War, called by Hans-Jurgen Bieling Keynesian period, when European regulatory framework (like Treaties of Rome) "primarily aimed at supporting national socioeconomic models and their development by providing an advantageous, growth and employment-friendly economic environment" (Bieling 2003, p.205). Then, in the 1970s and 1980s and with the programme of Single European Market, "Keynesian policies are being replaced step by step by a new, more aggressive configuration, which basically is neoliberal, i.e. in favour of broadened and intensified market competition and monetarist antiinflation and austerity measures" (Bieling 2003, p.206). This period is called the post-Fordist¹⁴ one, and also defines nowadays European Union.

Thus it becomes very simple for a theory of European integration to understand this process in a wider structure and link it with other worldwide transformations like globalisation. Furthermore, globalisation itself should not be understood as a distinct process from European integration, but as a complementary one. For Andrea Bieler and David Morton "globalisation can be defined as the transnationalisation of production and finance at the material level and the shift from Keynesianism to neoliberalism at the ideological level" (Bieler and Morton, p. 4). This process could also be found at European level by incorporating the new social forces "engendered by the transnationalisation of production and finance" (Bieler and Morton, p. 17).

To conclude, these social forces are in fact market forces like: lobby groups, banks, corporations or even unions. And, taking into consideration all this concepts,

⁹ It represents the formation and alignment of groups in relation to production. "This enables the examination of the question of whether *the necessary and sufficient conditions exist in a society for its transformation*"(Gill 2003, p. 51-52).

¹⁰ "This involves an assessment of the degree of homogeneity and political consciousness among different classes and political groupings" (Gill 2003, p. 52).

¹¹ "This primarily involves the relation of military forces, which from time to time is immediately decisive" (Gill 2003, p. 52).

¹² It was established by the Treaty of European Coal and Steel Community.

¹³ It was established by the Treaty of Maastricht.

¹⁴ The Fordist period is characterised by the volume production, while the post-Fordist period is characterised by the quality and specific production.

European integration process can be analysed using the relation between structure and super-structure¹⁵. The economic structure is understood here as being shaped by the social forces agency, while the ideological super-structure is characterised by the neo-liberal model.

3. Methodology

My research is a theoretical one, and I will analyse the works of the following authors for both parts of my paper: Andreas Bieler, Bastiaan van Apeldoorn, Peter Cocks, Adam David Morton, Stephen Gill or Robert Cox. Furthermore, the neo-Gramscian methodology which I will use in the analysis and critique of the classical and middle-range theories uses the *historical situation* as the object of the analysis, and the observation of force relations as the method of analysis (Gill 2003, p. 51).

4. Conclusions

Because the classical theories failed to explain the whole European integration process and middle-range theories already gave up on doing this, neo-Gramscianism takes a strong position in the puzzle of those theories. Being based on the neo-Marxist dialectical relationship between economic structure and ideological super-structure, neo-Gramscianism succeeds in understanding how some historical situations and a specific mode of production could determine changes in the evolution of the European Union.

Also, the first Marxist analyses of the European integration process were more focused on economic deterministic aspects and escaped with the super-structural aspects. With neo-Grmascianism theory, integration could be understood in a larger view, between the dialectical relation of the structure and super-structure. Thus, the neo-liberal dimension, determined by the structural arrangements, has also a deterministic role in the European policy output, and shape (due its hegemonic character) the ways of further European development.

It becomes very simple now to apply Cox's distinction between problem-solving theory and critical theory. Cox states that theories are created for someone and with a purpose and defines them as following: problem-solving theories legitimate the status quo of the reality they should explain, while critical theories are trying to deal with the system or structure that produces a certain reality (Cox 1986, p. 207-208).

References

Apeldoorn, B. (2002) "Transnational Capitalism and the Struggle over European Integration", Routledge, London.

Apeldoorn, B., Overbeek, H. and Ryner, M. (2003) "Theories of European Integration: A Critique" in Cafruny, A.W. and Ryner, M. "A Ruined Fortress?

¹⁵ In classical Marxism, structure determines super-structure. But many neo-Marxists show that also the super-structure can influence many changes in the economic structure of the society.

Neoliberal Hegemony and Transformation in Europe", Rowman & Littlefield Publishers, New York, pp. 17-47.

Bieler, A. and Morton, A.D. (2001), "Introduction: Neo-Gramscian Perspectives in International Political Economy and the Relevance to European Integration" in Bieler, A. and Morton, A.D. "Social Forces in the Making of the New Europe. The Restructuring of European Social Relation in the Global Political Economy". Palgrave, Basingstoke.

Bieler A. (2000) "Globalisation and Enlargement of the European Union: Austrian and Swedish social forces in the struggle over membership", Routledge, New York.

Bieling, H.J. (2003) "Social Forces in the Making of the New European Economy: The Case of Financial Market Integration". *New Political Economy*, Vol. 8, No. 2, pp. 203-224.

Cini, M. (2010) "Intergovernmentalism" in Cini, M. and Borragan, N.P.S., *"European Union Politics*", Oxford University Press, New York, pp. 86-104.

Cocks, P. (1980) "Toward a Marxist Theory of European Integration", *International Organisation*, Vol. 34, No. 1, pp.1-40.

Cox, R. (1986) "Social Forces, States and World Orders: Beyond International Relations Theory" in Robert o. Keohane, *"Neorealism and its Critics*", Columbia University Press, New York, pp. 204-255.

Gill, S. (2003) "A Neo-Gramscian Approach to European Integration" in Cafruny, A.W. and Ryner M. "*A Ruined Fortress? Neoliberal Hegemony and Transformation in Europe*", Rowman & Littlefield Publishers, New York, pp. 47-71.

Jensen, C.S. (2010) "Neo-functionalism" in Cini, M. and Borragan, N.P.S. "*European Union Politics*", Oxford University Press, New York, pp. 71-86.

Mandel, E. (1967) "International Capitalism and "Supra-Nationality"". Socialist Register 1967, pp. 27-41.

Morton, A.D. (2007) "Unravelling Gramsci: Hegemony and Passive Revolution in the Global Political Economy", Pluto Press, London.

Niemann, A. and Schmitter, P.C. (2009) "Neofunctionalism" in Wiener A. and Diez T., *"European Integration Theory*", Oxford University Press, New York, pp. 45-67.

Pollack, M. (2010) "Theorizing EU Policy-Making" in Wallace, H., Pollack, M. and Young, A. "*Policy-Making in the European Union*", Oxford University Press, New York, pp. 15-45.

Poulantzas, N. (2008) "The Poulantzas Reader", Verso, London.

Rosamond, B. (2010) "New Theories of European Integration" in Michelle Cini and Nieves Perez-Solorzano Borragan. "*European Union Politics*", Oxford University Press, New York, pp. 104-122.

Florenca Korbi, PhD Candidate¹,

University of Tirana, Department of History, Affiliation: International Relations E-mail: florenca_lleshaj@yahoo.com

Abstract.

Even though the subject of anti-Americanism has been addressed in the political literature in Europe and other regions of the world, it has sparked enough interest, especially after the tragic events of 11 September, 2001. However, until now a few things have been written about pro-American sentiments and attitudes in favor of the United States, which, of course exist.

The issue of values is one of the most controversial and most discussed topics in today's debate on foreign policy, as in the United States and in European Union member countries. There is no doubt that values themselves as well as their role on how to develop state policies, are of special importance. Also, the Europeans and Americans not only respect the same fundamental values, but sometimes fierce debate, about them. However there is still no clear concept that theoretically supports the existence of often contradictory views on the world of transatlantic relations.

The state of these relations and context in which international politics takes place today, makes the necessary elaboration of the anti-Americanism concept, by political science and international relations theory. Americanophobia and anti-Americanism are the strangest prejudices in modern history and even in theoretical doctrinal principles.

Anti-Americanism in Europe, in short, is a problem that emerges from European culture and it will be with us for a long time. Thus, the aim of this article is to understand its source, and how it is regularly remobilized to serve the needs of particular politicians. We also must be able to see anti-Americanism in Europe as an indicator of European culture, rather than the fault of U.S. policies. Historically, Europeans have been among the strongest friends of the United States, and for many that tradition continues but there is also, a European predisposition to vilify the United States. Then, what is the source of anti-Americanism? This paper elaborates it.

¹ Florenca Korbi, has finished graduate studies in 2007, at the University of Shkodra, Faculty of Foreign Languages, for English language. Her postgraduate studies cover the field of International Relations. Currently, she is attending doctoral studies in the same field, International Relations, at the University of Tirana.

Through literature review, I tend to explain how to understand anti-Americanism and Americano-phobia, the geographical distribution of anti and pro-American sentiments and above all, to portray the causes and consequences of the European anti-Americanism.

Keywords: Anti-Americanism, Americano-phobia, Europe, Popular Culture

Uunderstanding Anti - Americanism and Americano-Phobia

America, being what it is, do not leave anyone indifferent...It motivates great admiration, sympathy, gratitude, but also jealousy, frustration, need for protection but also need for independence.²

Taking into account the words of Hugo Paemen, I would say that the United States of America from their creation, have either loved or hated. These main trends were noticed: on one hand there were people who admired America, its people and its way of life, on the other hand, there were people who had worried the role that America has played and continues to play in the global scene.³

But what should we mean by anti-Americanism, is this a clear notion and well-defined? Does Americano-phobia, has the same impact as anti-Americanism? When I ask about this topic, I do not think that anti-American sentiments were so deep, so old, even that had arisen with the discovery of America. Reading Fatos Tarifa, Denis Lacorne, Hubert Vedrine, Philippe Roger, James W. Ceaser⁴ and many others, I found out that anti-American sentiments or earlier negative stereotypes about America, were fed in 1700s, by the pseudoscientific theory of Comte de Buffon who linked the degeneration and deficiency of the "New World" with it's very wet climate.

Precisely these completely irrational and absurd stereotypes have given life to Americano-phobia which in the view of the French author, Denis Lacorne, distinguished anti-Americanism. He said: "By anti-Americanism, I mean the critical and reasoned expression of a disagreement with what Americans say or do. By Americano-phobia, I mean the total visceral rejection of anything that has to do

² Paemen, Hugo.2003. Testimony on "The future Transatlantic Relations: A wiev from Europe" before the U.S House Committee on International Relations, Sub-Committee on Europe, 17 June

³ Tarifa, Fatos. 2008. Amerikanofobia dhe Anti-Amerikanizmi Europian, pg. 27

⁴ Ceaser, James W. 2003. *A Genealogy of anti-Americanism*, The Public Interest. (online source: <u>http://thepublicinterest.com/archives/2003summer/article1.html</u> accessed July 17, 2011.

with American culture, democracy, or economy, in short, with American civilization".⁵

David Ellwood, defines anti-Americanism as a cultural and political phenomenon which we can understand only by being familiar with its history, studying the evolution of all forms of antagonism toward the nation, people, civilization and the actions taken by United States of America. Meanwhile, for Fatos Tarifa, anti-Americanism is a negative predisposition, or a kind of ambiguity, which in most cases has not base objective or content. Anti-Americanism is a certain attitude.⁶

Personally I think that anti-Americanism and Americano-phobia reflect different reactions to the United States and which therefore have different impacts on the latter. While I see(or understand) anti-Americanism as a reflection of thoughts, opinions on the most critical cases, such as unfavorable judgments against the United States or its policies; I see Americano-phobia as a reflection of prejudice, as distortion of information predisposition to believe only the negative about the United States. At the same time, as long as I understand anti-Americanism as opinion, then there is the possibility that the latter be changed while the phobia against America, for me it's hard to influence. In this regard, Americano-phobia perceived as bias, misrepresentation of information, has major consequences on the long term on the United States, on their foreign policy, rather than anti-Americanism understood as a particular opinion.

However, things are not so simple. Anti-Americanism is a very vague concept for which there is no single definition. Moreover some authors talk about different types of anti-Americanism, for its variety forms. But let's get acquainted below with the history of anti-American sentiments in Europe, to understand better this phenomenon.

1. A Brief Historical Overview

European Anti-Americanism is a sentiment that has existed since the creation of America itself. Since then, European scholars have discussed about America and have even discoed without visited it a single time. Negative prejudices about America and Americans are not new. In an article written by James W. Ceaser "*A Genealogy of Anti-Americanism*"⁷, we can find a mosaic of negative stereotypes, paradoxical prejudices on the United States and the lives of Americans which have started since the mid- eighteenth century. Since at this time starts the whole history of anti-Americanism in Europe, a history which has adversely affected the transatlantic relationship, an absurd history but also harmful.

⁵ Lacorne, Denis. 2005. Anti-americanismand Americanophobia: A French Prespective, pg.14

⁶ Tarifa, Fatos.2008. Amerikanofobia dhe Anti-Amerikanizmi Europian, pg. 27.

⁷ Ceaser, James W. 2003. *A Genealogy of anti-Americanism*, The Public Interest.

⁽online source: <u>http://thepublicinterest.com/archives/2003summer/article1.html</u> accessed July 17, 2011.)

According to naturalists Buffon, Corneulius de Pouw and their "degeneration thesis", nobody could live in the "New World" whose discovery was the biggest tragedy that happened to mankind.⁸Their absurd prejudices, adversely affected the subsequent European intellectuals dominating in most of the XIX century. During this century, European aristocracy hated America because it had become a symbol

of progress, innovation and above all equality of opportunity. Precisely when this new nation created the American model of populist democracy increase mass culture, and spread industrialization, then Europe began to see America as a viral threat that could destabilize the established order in the old continent. Already, America accused of what would call excessive civilization, which will further feed Americano-phobia and anti-Americanism. As Fatos Tarifa notes, European anti-Americanism of this period, has been generally and consistently associated with social, cultural and political conservatism, with efforts to preserve old traditions and the protest against modernity.⁹

After the Second World War until the end of the Cold War American domination became less abstract. At this time is supposed that America was dominating the world and should have done do so. American foreign policy, the pressure of modernization, as well as being dollar empire, returned to the feeding force of anti-Americanism. An obvious fear of this hegemonic began to distort Europe, for over France during the first decade after the war. Mainly, European leftists mostly fed negative sentiments towards the United States which ranged from dislike up to irritation or antipathy. At this time, anti-Americanism became the hallmark of official policy of France which was feeling the fear of losing its cultural supremacy that it has enjoyed since that time.. Anti-Americanism of this period especially in '70 years, was more oriented to American culture, only later was supplemented by the antipathy to American state and policy. In short fear over this so powerful country that was changing all the balances that existed since that time, was modernizing everything, that was dominating everywhere, fueled anti-American sentiments in postwar Europe.

Later, with the collapse of the Soviet Union and the end of the Cold War, America became the sole superpower, being simultaneously the object of contempt and criticism, of disgust and even the envy of Europe. What is interesting relates to the paradox that has always accompanied America, at the same time this country has been the most offended and the most coveted; and"samples" and the desired model. Anti-American sentiments in Europe were strengthened at the precise time when the opposite was expected. Before 1989 America was the "protective umbrella" of Western Europe but after the collapse of the Soviet Union, the United States became undesirable in Europe and especially in France.

⁸ Diner, Dan. 1996. America in the Eyes of Germans, pg. 3-4.

⁹ Tarifa, Fatos.2008. Amerikanofobia dhe Anti-Amerikanizmi Europian, pg. 38

America began to be seen as damaging and dominant superpower, as a towering model of civilization, as a diseases collection of westernising and globalization.¹⁰ Anti-Americanism of this period was often connect with global movements, with the fear of world domination by the capitalist interests and American culture.¹¹

After the events of 11 September 2001, Western Europe's antipathy toward the United States, leaves the turn to solidarity for the tragic events that shocked not only America. But soon, this solidarity faded to be replaced by old systemic disease of Europe, anti-Americanism. In many aspects, terrorist attacks of September 11, were fundamentally anti-American acts, which meet all the definitions of anti-Americanism. Therefore, not in vain Theodore Zeldin has defined anti-Americanism as a pathology which explains these events. He wrote: To hate a whole nation, to love a whole nation, is a clear syndrome of hysteria.¹²

At the time of the outbreak of war in Iraq, it was born a new wave of anti-Americanism, which resulted mainly from strong feelings against the war that had some European countries. Western Europe, especially France and Germany, were inclined to see America as selfish and irresponsible, which according to them was risking the country to strengthen its Security Foreign policy pursued at this period by the W. Bush administration, led to the creation of rather than a negative image for America to European countries.

While today, the strengthen role of the United States as sole superpower in the world has risen and spread cultural, socio-economic and political anti-Americanism in Europe. The gap between Europeans and Americans has grown so much as France and Germany, are in charge of exporting countries for the anti-American sentiments. But how are spread anti-American sentiments in Europe, ultimately, are all Europeans anti-American? Let 's handle this issue below

2. The Spread of Anti-American Sentiments in Europe

All the kinds of anti-Americanism, are played with various melodies and rhythm in all states of Europe \dots ¹³

Not only Peter J. Katzenstein and Robert O. Keohane speak with such an argument but Andrei Markovits asserts that all forms of anti-Americanism as a naturalist, cultural, socio-economic and political, can be found throughout Europe.

¹⁰ Rubin, Judy C. 2004. *The Five Stages of anti-Americanism*. Foreign Policy Research Institute.(online source: <u>www.fpri.org/enotes/20040904.americawar.colprubin.5stagesantiamericanism.html</u> accessed July 11, 2011)

¹¹ Barber, Benjamin. 1996. Jihad VS. McWorld.

¹² Barber, Benjamin. 1996. Jihad VS. McWorld.

¹³ Katzenstein, Peter J. Keohane, Robert O. 2007. *Anti-Americanisms in World Politics*, pg. 57

This author, quoting Sigmund Freud, according to which"Yes, America is gigantic, but a gigantic mistake", and Charles Dickens, mocking the notion that America should arouse envy throughout the world, ¹⁴argues that what is at play is in reality a pan-European discourse of anti-Americanism, rather than a country specific one,¹⁵ in terms of its form and content, as well as its analytical foundation, concepts, and political and social functions .¹⁶ I think that this argument applied by Andrei worth to be taken into account to some extent, but not without Markovits reservations. First of all, it is clear that while anti-American sentiments are widespread around the world, they remain vulnerable in Europe more than anywhere else, prevalent in Western Europe, more than in the Islamic world and as I would like to point out, still more in France, which is defined as the epicenter of European anti-Americanism elits .¹⁷ But the problem is that while for this author, anti-americanism is a pan-European phenomenon, for other authors as Hugo Paemen, in Europe there is no general anti-Americanism. Even if we speak with the words of the former Polish Defense Minister Radek Sikorski, cited by Fatos Tarifa, in some European countries like France, anti-Americanism is endemic, widespread, while in other countries, for example in Poland, it almost does not exist.¹⁸ But what can we say about anti-Americanism in Germany, Britain or other European countries? Meanwhile, French anti-Americanism is not understandable, considering the fact that France has never been at war with the United States, by contrast, have been allied, the anti-American sentiments of Germany, Russia, Spain or England, I think can be somewhat acceptable or justified.

According to Russell Barmann the case of Germany is such that anti-Americanism and filo-Americanism exist side by side. While it shares common values with Americans, she can nurture anti-Western and anti-American subculture, as happened after 11 September 2001.¹⁹ Personally, I see the German and French attitudes toward the United States, not only in terms of direct assessment of the policy of the latter, whether past or future, but as a consequence of the assessment that the Germans and French have for the their role in world affairs. Today's unipolar system, has left them a space in which neither Germany nor France are satisfied.

Of course, not all Europeans are anti-American, but we must be aware that in certain circumstances, some European politicians choose to play the "*anti-Americanism card*" where they are more expedient, especially in the circumstances of their election. As happened with Jasques Chirac and Gerhard Schröder in Germany in

¹⁴ Markovits, Andrei S. 2007. Uncouth Nation. Why Europe Dislikes America, pg. 56-73

¹⁵ ibid, pg. 28

¹⁶ ibid, pg. 46.

¹⁷ Katzenstein, Peter J. Keohane, Robert O. 2007. *Anti-Americanisms in World Politics*, pg. 36

¹⁸ Sikorski, Radek. 2003.Testimony at a hearing session on "The Future of Transatlantic Relations: A view of Europe" before the U.S House Committeeon International Relations, Sub-committee on Europe, June 17, in Tarifa, Fatos. 2008. *Amerikanofobia dhe anti-Amerikanizmi Europian*, pg. 77

¹⁹ Barman, Russell. A. 2004. Anti-americanism in Europe: A Cultural Problem, pg. 34

2002, who by following this strategy reached to win the elections. Gerhard Schröder,German Chancellor accused the only world's superpower, that it was not an "organized state." In 2002, he built his election campaign, criticizing the wrong policies of the United States, especially the Administration of the President George W. Bush. He fielded himself as anti-American supporting the President Jacques Chirac, and the President Vladimir Putin to counter American hegemony.

However, Europe has historically enjoyed close relations with the United States, a relationship that no other part of the world has had such at that level. The whole concept of "New World" versus "Old World", despite the common values they share, as Richard Pells notes, originated in Europe, and as such is carved in the form of deep convictions European anti- Americanism²⁰, making that modernity seen as barbarism, to be a negative and constant perception of Europe to America.²¹

What we should consider when we talk about Europe, is that we should not perceive it as a single whole homogeneous. European countries represent cultural diversity and cultural, historical and geographical differences. Furthermore, as Fatos Tarifa stated, the similarities in appearance between the European Union, in fact conceal significant cultural differences and quite obvious national identities. He is right in thinking that Britain is culturally very similar to the United States than to the European countries like France, Italy, Greece or Poland.²² However, although it is unclear what they have in common, it is argued that one of the few values that all Europeans share in the moment made against America, is "*not being American*."²³ American sentiments, which in various forms have permeated all these countries, have helped Europe to create a self identity against America.

Negative perception that Europeans have against cultural, political and economic influence, or security issues in their countries, the rejection to American power, I think that comes from the need of Europe to be more independent in all these areas, either in matters of security,economic issues or diplomatic ones. But on the other side, feeling not ready for this independence, proves at the same time a confrontation of sentiments that revolve around the need to be under American protection at one hand and the need to be independent at the other. Writing these words, I bring to mind Philippe Roger which says, *is the feeling of failure and disappointment what makes French, hostile and aggressive*,²⁴... maybe this applies to all anti-American Europe.

After we talked so much about European anti-Americanism, may naturally ask, are all Europeans anti-American? Are all Europeans aggressive and hostile to foreign policy, socio-economic system and above all to the American culture that is widely spread in the "era of globalization"? Of course not. It is true that much has been written about anti-American sentiments in the European continent and especially

²⁰ Pells, Richard.1997. Not Like Us. How Europeans Have Loved, Hated, and Transformed American Culture Since World War II, pg. 2-7

²¹ Markovits, Andrei S. 2007. Uncouth Nation. Why Europe Dislikes America, pg. 46

²² Tarifa, Fatos.2008. Amerikanofobia dhe Anti-amerikanizmi Europian, pg. 23

²³ Ash, Timothy Garton. 2004. *Free World. Why a crisis of the West reveals the opportunity of our time*, pg. 54-94

²⁴ Roger, Philippe.2005. *The American Enemy: The History of French Antiamericanism.*

in Western Europe, on the grievance, hatred against America, while very little is spoken about pro-American feelings in Europe, but this does not mean that they not exist.

In fact, the relationship between the United States and the countries of Central and Eastern Europe by several authors, is defined as "*special*", meaning here a special and different relationship not just from what the relationships of the United States have been in similar regions but also the relationship that these small countries have historically been with the major powers. In the case of Albania "special relationship" remains valid even after Albania appears to be more pro-American nation in the world, followed by Kosovo Albanians. ²⁵And in fact, that in the early 90s had clear signs that the foreign policy of this country was moving to align with the United States which had strongly supported the democratization of Albania. Not only in the early 90s but today perhaps there is no country in Europe, to have such sympathy for the United States as Albania.

However, in most Central European countries, mainly Eastern Europe and Eastern countries of the former communist bloc, the United States are seen as "beacons" of their freedom and democracy. American presence in Europe has always been considered necessary counterweight against the dominant trend of France or Germany. In contrast to Western European, the concern of American cultural domination was not present in Eastern European countries which for many years had to face the Soviet cultural domination and the communist regime. Moreover, American culture is always seen as attractive in these countries. Also in contrast to Western European, the countries of this region seem genuinely grateful to the United States for their freedom, so not in vain are defined by Donald Rumsfeld, former Secretary of Defense of United States, as "*New Europe*" pro-American.²⁶

3. Causes and Consequences of the European Anti-Americanism

After the Second World War, relationships between Europe and America, have undergone constant fluctuation, but the size and strength of antipathy towards America today in Europe, exceed and they of the most tense period in relations between the two sides of Atlantic. In the last years, all of us have witnessed the current situation between the Europe and United States. In this situation, anti-Americanism should be see just one event, a cause, or a result in this transatlantic disruption?

Anti-American sentiments in Europe, are not only due of the foreign policy of the United States, of their hegemony, military force, or only result of "hatred" of these countries towards Anglo-Saxon socio-economic model that Europe in general has refused to imitate. Andrew Kohut, based on the data of "Pew Global Attitudes

²⁵ Tarifa, Fatos. 2008. Amerikanofobia dhe Anti-Amerikanizmi Evropian, pg. 117

²⁶ A transcript prepared by: Federal News Service INC., Washington, D.C. Secretary Rumsfeld Briefs at the Foreign Press Center. News Transcript. January 22, 2003

Project," found four aspects, under which are the central reasons that have led to grow anti-Americanism worldwide and specifically in Europe:²⁷

- 1. A general perception that the U.S. acts unilaterally in the international arena, failing to take into account the interests of other countries when it makes foreign policy decisions.
- 2. A broad discomfort with unrivaled American power.
- 3. A perceived disproportionate willingness to use military force, and especially preemptive force.
- 4. A rejection of "Americanization" the wide diffusion of American ideas and customs fueled by globalization.

Recognizing the fact that the causes which have led on the rise of European anti-Americanism are numerous, I would like to stop at those causes which Joseph S. Nye has designated as sources of "*soft power*"²⁸ in America and include among others, *political values and cultural norms*.

As we know, the United States historically have fought to spread their values throughout the world being somehow self-proclaimed as the universal nation, whose values should be applied universally to all mankind.²⁹ But such an attitude can not pass without challenge and without becoming subject to the criticism of those nations that historically have "competed for the title" having the same aspirate: exporting their values to the rest of the world. From this perspective, it is understandable why the European national creed is colliding with the U.S. national creed, it is understandable why for France has been so difficult to accept the spread of American values throughout the world, it is understandable why European countries feel the need to protect their values and culture and is therefore clear why anti-Americanism is fueled by the so-called "imperialism" of cultural and American values

Indeed, the idea that America is different from the other countries and belongs the mission of preaching its values in the rest of the world, can be found in all sides of American politics and has been a source of influence on foreign policy of this state. The words of Secretary of State Madeleine Albright, according to which "*the American nation is necessary*", clearly point to elements of American national identity. Also, through these words we can understand why many Americans are surprised when confronted with anti-American sentiments. The experience of becoming a model for the rest of the world, lies deep in their national consciousness, even claimed that they necessarily require that other nations adopt the same values and virtues of the great nation of the future. All this can be best illustrated by the words of James Skillen, who writes:

²⁷ Kohut, Andrew.2007. America's Image in the World: Findings from the Pew Global Attitudes Project. Testimony of Andrew Kohut, President of the Pew Research Center, to the Subcommittee on International Organizations, Human Rights, and Oversight Committee on Foreign Affairs, U.S. House of Representatives, March 14. pg. 5-7.

²⁸ Nye, Joseph S. 2004. *The Decline of America's Soft Power*. Foreign Affairs, Vol. 83 Issue 3.

²⁹ Lieven, Anatol. 2005. America Right or Wrong. An Anatomy of American Nationalism, pg. 34

"Americans have believed from the beginning that their nation has a divine mission to fulfill, to bring light to the world, a light that was subsequently defined as the light of freedom, democracy, and prosperity rather than the light of biblical righteousness." ³⁰

The problem lies in the fact that this American national mission is contrary to national interests of anti-American nations in Europe, that not only don't like to be guided by the United States by this "dominant, obnoxious allied, that always act as triumphant, selfish and hegemonic superpower" but they criticize and denounce as controversial the most part of what those latter perceive as values. On this point of view, would be entitled Markovic who emphasizes that the European anti-Americanism relates more to America's identity, thus *what America is* than its behavior, thus *what America does*.³¹ In this context, the typical American values, and so more the ones they are trying to export through their foreign policy, have served as a repeated source of European anti-Americanism.

Regarding the issue of cultural norms, we can say that the United States have become a synonymous with the so-called *cultural modernity*: American cultural products and norms, have become global in a much greater extent than those of the other countries. As Lieber said, while priority and influence of the United States are easier to assess in the economic and military sphere, in the cultural arena, it becomes more difficult, however, even here, American superiority is really noticeable.³²

In fact, the cultural dimension of the United .States "soft power" is very important because the United States, are increasingly seen as the epitome of mass culture. American popular culture, where we can talk about fast food industry popularly known as the American fast-food, pop music, television, Internet pages indicated as computer-based media, constitute the most widespread and prominent exports American, which first became attractive to European people, later to the entire world. Exactly this process has frightened European political and cultural elites, according to which American mass culture poses a serious threat to the "top" culture of the old continent, massive imports of American cultural products are degenerating their own culture. Seen in this light by the European cultural elites, cultural anti-Americanism is growing day by day. This form of anti-Americanism is fed constantly being manifested in the form of rejection of American cultural products on the one hand and constantly criticizing their attractiveness to the peoples of the world, in the other.

However, except the American unilateralism as alleged in the post Cold War world, except the American primacy in everything as a single global hegemony in today's uni - polar system, except the clash of political values, globalization and so

³⁰ Skillen, James W. 2005. With or Against the World? America's Role among the Nations, pg. 78

³¹ Skillen, James W. 2005. With or Against the World? America's Role among the Nations, pg. 78

³² Lieber, Robert J. 2005. *The American Era. Power and Strategy for the 21st Century*, pg. 101

on, the causes of European anti-Americanism are constantly fluctuating between historical or current European countries identity and cultural norms confrontation between the two continents, making European anti-Americanism widely perceived as a problem of culture.

Conclusions

Finally we can say that the close relations between Europe and the United States, today we find really in low level. European anti-Americanism has created tensions and more difficulties to transatlantic relationships. American dominance has generated frustration among those who feel threatened and scared by it and the Europeans are feeling each day and more the need of protecting their culture from Americanization, attempting to resist to the universal tendency of this phenomenon. However, Europeans and Americans should realize that they cherish and share common values on the basis of which, should continue to cooperate and provide solutions to global problems.

European anti-Americanism, especially the cultural is not a new phenomenon, two sides of the Atlantic have lived for more than half a century with this problem: the cultural division between a vigorous and masculine America and a sophisticated, feminine Europe, without backbone. However, Europe and America could manage better these differences, as they could afford and the political tension caused between them.

Absolute superiority of the United States in global affairs, especially if this phenomenon is seen in relation to a relative decline of Europe, their unilateral foreign policy, their hegemony, military force, the "hatred" of European countries towards Anglo-Saxon socio-economic model but even more the *cultural confrontation between the two continents*, remain the main causes that have inflamed and intensified anti-American sentiments in many corridors of European diplomacy and politics, particularly in France, making anti-Americanism almost a global phenomenon.

But what I can say is that anti-Americanism in Europe has affected the transatlantic alliance and if so extended, will only further encourage Americans to give unilateral answers to common global problems. According to my opinion, is the duty of the European states, of the European moderators to refuse and to give up from the basic rhetoric of anti-Americanism, which is harmful in nature and free by any form of rationality.

References

- 1. Ash, Timothy Garton. 2004. Free World. Why a crisis of the West reveals the opportunity of our time. London: Allen Lane.
- 2. Barman, Russell. A. 2004. *Anti-americanism in Europe: A Cultural Problem.* Hoover Institution Press. Stanford University.
- 3. Barber, Benjamin. 1996. *Jihad VS. McWorld*, Ballantine Books, New York.
- 4. Ceaser, James W. 2003. *A Genealogy of anti-Americanism*, The Public Interest. (online source: <u>http://thepublicinterest.com/archives/2003summer/article1.html accessed July 17, 2011</u>.)
- 5. Diner, Dan. 1996. *America in the Eyes of Germans*, Princeton: Markus Wiener Publishers.
- 6. Katzenstein, Peter J. Keohane, Robert O. 2007. *Anti-Americanisms in World Politics*. Ithaca, NY: Cornell University Press.
- 7. Kohut, Andrew. 2007. *America's Image in the World: Findings from the Pew Global Attitudes Project.* Testimony of Andrew Kohut, President of the Pew Research Center, to the Subcommittee on International Organizations, Human Rights, and Oversight Committee on Foreign Affairs, U.S. House of Representatives, March 14, 2007.
- 8. Lacorne, Denis. 2005. Anti-americanismand Americanophobia: A French Prespective. (online source: http://www.cerisciencespo.com/archive/mars05/artdl.pdf-accessed August 7, 2011.)
- 9. Lacorne, Denis. Jacques Rupnik & Marie-France Toinet. 1990. *The Rise and Fall of Anti-americanism: A Century of French Perception*. St. Martin's.
- 10. Lieber, Robert J. 2005. *The American Era. Power and Strategy for the 21st Century*. New York: Cambridge University Press.
- 11. Lieven, Anatol. 2005. America Right or Wrong. An Anatomy of American Nationalism. London: Harper Perennial.
- 12. Markovits, Andrei S. 2007. *Uncouth Nation. Why Europe Dislikes America*. Princeton, NJ: Princeton University Press.
- 13. Nye, Joseph S. 2004. *The Decline of America's Soft Power*. Foreign Affairs, Vol. 83 Issue 3.

- 14. Paemen, Hugo. 2003. Testimony on "The future Transatlantic Relations: A wiev from Europe" before the U.S House Committee on International Relations, Sub-Committee on Europe, 17 June.
- 15. Pells, Richard.1997. Not Like Us. How Europeans Have Loved, Hated, and Transformed American Culture Since World War II. New York: Basic Books.
- 16. Roger, Philippe. 2005. *The American Enemy: The History of French Antiamericanism.* Chicago: The university of Chicago Press.
- 17. Rubin, Judy C. 2004. *The Five Stages of anti-Americanism*. Foreign Policy Research Institute.
 (online source: www.fpri.org/enotes/20040904.americawar.colprubin.5stagesantiamericanism.h
 <u>tml</u> accessed July 11, 2011.)
- 18. Sikorski, Radek. 2003. Testimony at a hearing session on "The Future of Transatlantic Relations: A view of Europe" before the U.S House Committeeon International Relations, Sub-committee on Europe, June 17.
- 19. Skillen, James W. 2005. *With or Against the World? America's Role among the Nations*. Lanham: Rowman & Littlefield Publishers.
- 20. Tarifa, Fatos. 2008. *Amerikanofobia dhe Anti-Amerikanizmi Europian*. Tiranë: Ombra GVG.

Representations of the Iraq war in two doctrines: neoconservatism and interventionism Oana Elena Brânda¹,

¹ History Faculty, University of Bucharest, oana.branda@gmail.com

Abstract. The debut of the March 2003 Iraq war has been surrounded by controversy, since its root causes go back in time to several years before. It is common belief that the events of 9/11 lie at the basis of the military interventions in Afghanistan and Iraq. And it is a true fact that after 9/11 the decision-making units in London and Washington constanly referred to the Middle East, namely Aghanistan and Iraq, as issues that need to be tackled as soon as possible.

The case for military intervention in the Middle East rests on the unfulfilled UNSC resolutions dating back to the first Gulf War. However, the lack of compliance with the UN inspections and the constant violations of human rights led the US and UK, after the tragic events of 9/11, to believe it was high time to intervene in the Middle East and democratize it.

Our aim is to analyze the manner in which the Iraq war is viewed through the prism of the two doctrines. We shall not focus on the events during and after the war, since our area of interest is centered around the debates that led to the common intervention in Iraq in 2003. We intend to offer examples of the two doctrines at work during the debates concerning the necessity and the manner to pursue such a war. What is more, we will analyse the manner in which decision-making vis-a-vis the Iraq war was influenced by the precepts of the two doctrines.

Keywords: Iraq war, neoconservatism, interventionism, terrorism, decision-making

1 Introduction

The debut of the Iraq war of 2003 has been highly controversial, due to the lack of concrete United Nations legitimation in this regard. The main supporters of a military intervention in Iraq with the aim to rid the world of the dangers of terrorism, weapons of mass destruction and the evilness¹ of Saddam Hussein, were the United States and the United Kingdom. In light of the terrorist attacks of 9/11 which not only struck the United States of America but also destabilized the international community, something had to be done to eradicate these threats. "Terrorism" and "weapons of mass destruction" have become terms often employed by the leaders of the USA and UK in their speeches concerning security, and this is especially visible after 9/11. The Iraq intervention of 2003 is considered a natural consequence of the terrorist attacks and the normal course of action to be pursued in order to protect the international community from further spread of this evil.

Although the United States and the United Kingdom have not been the only countries to involve in the Iraq war, their positions concerning this matter have been the most visible. It must be specified from the beginning of this essay that what lies at the foundation of these countries's responses to the terror threat, is in fact a very theoretical basis. It is a fact that the USA and the UK had different approaches to the Iraq war and these two different approaches stem from the doctrines manning the decision-making units – the neoconservatism of the Bush administration, molded rapidly as a reaction to the terrorist attacks of 2001, and the doctrine of the international community of the Blair Cabinet, formulated in 1999, focusing on the need to intervene in those countries where human rights and democracy are threatened, a doctrine also referred to as interventionism.

These two doctrines have generated two different types of approaches that have been translated into practice before and during the war. The manner in which these two doctrines have been employed to justify the route to war has produced a series of debates that suggest that such arguments may not have been the most suitable for the respective cause.

The aim of this paper is firstly to present the two doctrines individually and then to analyze the manner in which the Iraq war is viewed through the prism of the two doctrines. Our attention shall not be focused on the events during and after the war, but rather on the way in which the two doctrines influenced the course of events that led to the war. We shall deliver in the first place a theoretical view of the two trends and then will attempt to see how the events of March 2003 were molded in place by the two doctrines.

Before going further we must draw to the readers 'attention the fact that these doctrines are visible especially in the discourse of the two administrations. As a result, the present study shall focus more likely on the rhetoric inside the doctrine rather than on the doctrine *en ensemble* as a factor that led to the involvement of the two countries in Iraq.

¹ We have taken the right to use the terminology already employed by President Bush when referring to the states that pose real dangers to democracy – see here the "axis of evil"- Iran, Iraq and North Korea;

2. Literature review

An extensive literature has been written on the topic of neoconservatism in the United States and the New Labour that inspired neoliberalism in the United Kingdom. Off all the literature consulted in the field, Fukuyama's book *America at a crossroads* and Maria Ryan's *Neoconservatism and the New American Century*, are representative works on the evolution of neoconservatism from the 19th century Protestantism to the present day political rather than religious current. As far as British neoliberalism is concerned, only New Labour benefitted from large research, while neoliberalism as such has only been referred to as a natural result of the rebranding of the Labour Party in mis-90s.

One must bear in mind the fact that such literature bears the bias of either theorizing, or being too critical, but never realistic about the topic. As a result, the view expressed in the present article on the two currents/doctrines/trends is the result of reading all available materials and extracting the basic features that recommend the two as noteworthy representatives of the analyzed countries.

It is a fact that the all the references are of Anglo-American nature. Taking this into consideration the unilaterality of bibliographical resources, keeping the analysis rather moderate as far as the critical view is concerned has been quite a challenge.

3. Methodology

As far as methodology is concerned, the present work combines two main approaches: that of international relations and that of political sciences. This article was is intended to be analytical but not critical. We shall begin with the theoretical fundaments of the two doctrines and the end with *la mise en place* of them during the Iraq war of 2003.

There are also other methods that will become visible along the paper, such as that of the chronological discourse or the geopolitical analysis. The choice of these methods is given by the necessity to offer both a closer view of the events, in their individuality, as well as a larger frame of the context and background in which events occurred and doctrines were enacted and put into practice.

4. Analysis

4.1. Blair's Doctrine of the international community/Interventionism

The doctrine of the international community has its roots in the speech Prime-minister Tony Blair delivered in front of the Chicago Economic Club on the 22nd of April 1999 and in two trends of international relations previously employed by the foreign policy of the United Kingdom – internationalism and atlanticism.

Internationalism represents a belief in the importance of full global participation to the security and prosperity of one's own nation and all nations as well. Internationalists favor active diplomacy and vigorous international trade and cooperation. Internationalism has been a constant topic brought up by the Labour Party in convening upon Britain's global position.

On the other hand, atlanticism is a doctrine of transatlantic cooperation: a doctrine assuming that both western Europe and the United States can benefit politically and economically from cooperation, especially in military matters. Atlanticism is what has been shaping British foreign policy after 9/11, but this aspect shall be discussed later.

Combining atlanticism and internationalism was in Prime-minister Blair's view the key to turning the United Kingdom into a global actor, with a very developed moral side.

In his speech of April 1999 in Chicago at the time when the Kosovo crisis was unfolding, Prime-minister Tony Blair outlined the main principles of the new doctrine of the international community: reforming the international financial system, boosting free trade inside the World Trade Organization, reconsidering the UN framework, re-examine NATO, further coperation in meeting the targets of the Kyoto protocol and solving the issue of Third World debt.

Blair's particular contribution to a new doctrine of international security is framing security within 'globalisation', as an aspect of 'globalisation' alongside the more familiar economic and political aspects ². Blair's speech of 1999 divides the international community into "us" and "them" - "they" are the ones who terrorize the global community. It is here that Blair expands the concepts of "community" and "partnership" from a national to an international level. Eventually, it is all about re-imagining the "international community".

Blair's Chicago speech articulated a set of five criteria that need to be fulfilled in order to justify a military intervention. Firstly, a degree of certainty regarding the case in question must be established. Secondly, all diplomatic options need to be exhausted. Thirdly, the dilemma stands with the possibility of undertaking military operations for the respective case. Fourthly, a long term involvement is considered. Finally, the question of national interests at stake arises.

Unlike the Bush doctrine that we shall discuss later in this paper, the Blair doctrine was seeking for a justification for using force and not a reason. The backbone of this controversial doctrine (at least as far as the British are concerned) stands on global security and the protection of universal values that all democratic countries have embraced. Blair does not refer to a specific threat that menaces the international community, as was the case three years later in the Bush doctrine, but he refers more likely to the possibility of worldwide acknowledged values to be challenged. The roots of the Blair doctrine lie in the precepts of liberal interventionism.

The entry into force of the doctrine of the international community has strong connections with the Third Way policy employed by the Labour Party under Blair, envisioning a society having rights and responsibilities.

During the unfolding of the Kosovo crisis, Prime-minister Blair was concerned with the United States adopting a rather isolationist position, eager not to interfere any more in worldwide conflicts. This was more acute since at the time, it was Blair's decision to turn the Americans into veritable internationalists.

² Norman Fairclough, *Blair's contribution to elaborating a new 'doctrine of international community'*, in Lilie Chouliaraki(ed), *The soft power of war*, Amsterdam, John Benjamins, 2007;

952

Blair's doctrine was to be applied to international security and humanitarian intervention. It is based on invoking the idea of a just war, not in order to achieve territories, but to preserve global values.

Tony Blair calls for a new basis for international law, based on "global rules" which gives states a "duty and a right to prevent the threat materialising; and... a responsibility to act when a nation's people are subjected to a regime such as Saddam's" ³. It is here that Blair calls for the modification of the international norms to provide for an intervention not just for self-defence but also to correct injustices.

No matter the wide publicity the Chicago speech received, as well as the consequent Blair doctrine, the latter was never mentioned in any defence or security related documents issued by the Cabinet. This leads to the obvious assumption that the doctrine lacked any kind of influence. The Blair doctrine seems to have been completely ignored in Whitehall, which was not the case of the Bush doctrine and the State Department and Pentagon. These differences in practice reside in the power relations existing inside the British Cabinet and inside the American Government.

The fact that the British authorities failed to translate the Blair doctrine into a tangible document on security and defense can only mean that this doctrine encompasses a series of faults. Firstly, the doctrine is addressing the international community. However, it is very difficult to restrain this international community to some basic geographic coordinates that could help identify it. Moreover, the question of values arises: which are the representative values of the international community and who decided whether they are so good as to justify a military intervention to protect them. Thirdly, the aspect of this type of intervention heralded by Tony Blair suggests that such an intervention will occur only after the commencement of a negative event. For instance, in case of genocide, based on the doctrine of the international community, states will only interfere once they receive proof of genocide undergoing. Finally, the doctrine makes no reference to the involvement of the United Nations, which makes countries applying it liable to look forward to fulfilling their own interests throughout such a mission⁴.

Moreover, the Iraq war of 2003 contributed to the destruction of the Blair doctrine. Intervening in Iraq was not a humanitarian intervention since there are many authors who argue that Blair challenged immensely the norms of international law in order to be able to ensure a weak justification for intervening along with the United States there.

4.2.Neoconservatism

Neoconservatism as a concept is based on universalism and exceptionalism. It presents itself as a voice of the American people, expressing their values and beliefs that might be eroded from inside if people succumb to realist rhetorics.

The Neoconservatives were extremely virulent during the years of the Cold War. Once the threat posed by the Soviet Union disappeared, the Neocons lost

³ Doctrine of the International Community, Chicago Economic Club, April 22nd, 1999, available at <u>http://www.pbs.org/newshour/bb/international/jan-june99/blair_doctrine4-</u>23.html

⁴ Chriss Abbott refers to the many ambiguous elements of Blair's doctrine in *Rights and Responsibilities: The Dilemma of Humanitarian Intervention*, in *Global Dialogue*, Volume 7, Number 1–2, Winter/Spring 2005—*Humanitarian Intervention*;

strength in asserting their interventionist goals. However, in the last decade of the XXth century, the Neocons re-emerged with a different rhetoric. In light of the foreign policy led by the Clinton administration and the lack of strength of the Republican Party before the elections of 2000, the Neocons changed their discourse into arguing that the world was neither bipolar, nor multipolar, but rather unipolar⁵. Naturally, the United States were the beacon of this world, and this meant that they would be the decisive force in any conflict they might involve in. Furthermore, a change in strategy was necessary, from defensive to offensive in order to be able to preserve this unique position. The catalyst that served to activate this policy after the election of George W. Bush as President was the events of 9/11.

After the events of 9/11, neoconservatists believed that this current was the necessary trend to steer the country through the effective response to terrorism and weapons of mass destruction. As the Afghanistan and Iraq wars tended to show that the effective response was not that efficient, realists raised to surface with claims that neoconservatism should be replaced by realism⁶ and that the United States should reconsider their appetite for war and choose peace more likely.

An important feature of neoconservatism is its desire to export democracy⁷. This lies at the basis of its interventionist policies, as well as the emphasis put on the protection of moral values. The international fight for democracy which became a mantra for the US stems from this very doctrine. However, in the long run the doctrine is a far-fetched one. Turning arrogant regimes that might endanger peoples and even the global environment, into peaceful democratic countries might seem extremely simple at a mere glance. Nevertheless, in practice, it would require the total removal of that country with all its features and replacing it with a new one, bearing democratic markers.

The Neoconservatives do not accept a value-free view of the national interest as did the realists of Hans Morgenthau and Henry Kissinger. America needs to assert its interests as it only works for good purposes.

Moreover, the Neoconservatives consider the post-Cold War multilateral order as infringing upon their possibilities to exercise hegemonic power.

The way the Cold War ended shaped the thinking of supporters of the Iraq war, including younger neoconservatives like William Kristol and Robert Kagan⁸, in two ways. First, it seems to have created an expectation that all totalitarian regimes were hollow at the core and would crumble with a small push from outside. This over-optimism about postwar transitions to democracy helps explain the Bush

⁵ Maria Ryan, *Neoconservatism and the New American Century*. New York, Palgrave Macmillan, 2005, pg. 2;

⁶ Michael Williams and Brian C. Schmidt, , *The Bush Doctrine and the Iraq War: Neoconservatives vs. Realists*, paper presented by Michael C. Williams at the American University of Paris on Wednesday the 13th of February 2008, pg. 1;

⁷ Maria Ryan, *Neoconservatism and the New American Century*. New York, Palgrave Macmillan, 2005, pg.4;

⁸ These two are representative figures of neoconservatism and that is why we have decided to mention them more often in the present article;

administration's apparently incomprehensible failure to plan adequately for the insurgency that subsequently emerged in Iraq⁹.

Overestimation of the threat was then used to justify the elevation of preventive war to the centerpiece of a new security strategy, as well as a whole series of measures that infringed on civil liberties, from detention policy to domestic eavesdropping¹⁰.

"We need in the first instance to understand that promoting democracy and modernization in the Middle East is not a solution to the problem of jihadist terrorism; in all likelihood it will make the short-term problem worse, as we have seen in the case of the Palestinian election bringing Hamas to power. Radical Islamism is a byproduct of modernization itself, arising from the loss of identity that accompanies the transition to a modern, pluralist society. More democracy will mean more alienation, radicalization and — yes, unfortunately — terrorism"¹¹. This is the view of former Neoconservatist Francis Fukuyama, who rejected this doctrine short after its failure to deal effectively with the Iraq war situation.

It is a fact that nothing prepared the Bush administration for the war and its outcome. But in wars these things are common. In Bush's speech delivered immediately after the events of 9/11, the President spoke of the nation's duty to history and the need and duty to rid the world of tyranny.

And in case of a war, the Bush administration needed a doctrine to imbue all the precepts that would steer the nation along difficult times. This is how the Bush doctrine was born, through the framework of the National Security Strategy and disolving in itself a considerable number of neoconservatist elements.

If originally, the Bush doctrine was a strong supporter of war against the perpetrators of 9/11, it gradually evolved into including preemptive and preventive war against a series of regimes that might become enemies of the United States, at a later hour, even if the threat from these possible enemies is not imminent.

The Bush doctrine blended together idealism and realism, going back to the teachings of the Founding Fathers of the American nation. Abraham Lincoln and Woodrow Wilson are its sources of inspiration, who are also some of the sources of inspiration of the neoconservatives. But one of the mistakes of the neocons is the fact that even if they revendicate their doctrine from Woodrow Wilson, they do not confide in the role of international organizations of maintaining peace but rather in the need of a unilateral intervention of force.

The Bush doctrine and neoconservatism need not be mistaken for one and the same. What differentiates the two is the evangelical trend that characterizes Bush's speech. More likely, the Bush doctrine is imbued with several neoconservative elements.In order to make the doctrine work, one needed good foretelling of the future and great intellignce. As both did not happen and since the United States opted

⁹ Francis Fukuyama, After Neoconservatism, February 19, 2006 available at http://www.nytimes.com/2006/02/19/magazine/neo.html?pagewanted=all

¹⁰ We are referring here to the Federal laws regarding the tapings of telephone conversations. More pieces of information are available here <u>http://www.callcorder.com/phone-recording-law-america.htm;</u>

¹¹ Francis Fukuyama, After Neoconservatism, February 19, 2006 available at http://www.nytimes.com/2006/02/19/magazine/neo.html?pagewanted=all

for unilateralism, this led to a even higher degree of isolationism than in the 19th century when the Monroe doctrine was consecrated.

There are two American administration characters often associated with neoconservatism and the Bush doctrine. These are Donald Rumsfeld and Dick Cheney. Although the two are not theorists of neoconservatism such as William Kristol and Robert Kagan who crystallized the doctrine at the end of the XXth century, the former are responsible of having embedded neoconservatism in the Bush administration. They shared the belief that unipolarity was a feature of US foreign policy that needed to be transformed into a permanent state.

But how did the neoconservatives integrate themselves so well in the Bush administration? Moreover, how did they manage to blend the neoconservative theory with the need to act against Iraq and bring along regime change in the Middle East? The answer dates back to the time before George W. Bush became president.

Long before President George W. Bush took office in 2001, elements in or close to the Republican Party had called repeatedly for firmer U.S. steps against Iraq, including a war if necessary to force a regime change. One such group authored a white paper in 1996 called "A Clean Break: A New Strategy for Securing the Realm", which was later sent to Prime Minister Benjamin Netanyahu, leader of Israel's Likud Party. It advocated a war against Iraq as a way of undermining Syria and of moderating the Shia Hezbollah of southern Lebanon, arguing that these actions would pave the way for peace and stability in a notoriously unstable part of the world. The paper came out of discussions among foreign policy experts, including Richard Perle, Douglas Feith, Robert Loewenberg, David Wurmser, and Meyrav Wurmser, many of whom later occupied important positions in the Bush administration.

In 1997 some of the same individuals joined the newly formed Project for a New American Century (PNAC), a Washington think tank that argued openly for the United States to play a dominant role, militarily and diplomatically, in the world. Faced with a more peaceful approach on the matter from the Clinton administration, a strong advocate of containment in the region, the group had to stand back and wait for more favourable times to come.

And these times came along with the entry to the White House of George W. Bush. It was now the time for furious neoconservatists such as Richard Perle, Paul Wolfowitz, Douglas Feith, Elliott Abrams to take a stand and make a point on the topic of the Middle East.. They, as well as others in their circle maintained that by democratizing Middle East countries with authoritarian regimes, the chances were greater of promoting peace in that region. In addition, many of these advisers were politically sympathetic both to the right wing of the Republican Party and to the Likud Party in Israel. Many had been, or their parents had been, on the political left, but they had typically become Republicans in the late 1970s or in the 1980s, driven by a belief that the Democratic Party was soft on the Union of Soviet Socialist Republics (USSR) and that the American left was increasingly sympathetic to the Palestinians at the expense of Israel. Because of their turn to the right, they were known as neoconservatives.

Because of their key positions in the Department of Defense, including in Vice President Dick Cheney's own national security council, and in the Near East and South Asia division's Office of Special Plans under Feith, the neoconservatives were in a position to influence Bush administration policy on Iraq. Some critics accused them of being overly eager to believe shaky intelligence on alleged Iraqi weapons of mass destruction programs provided by expatriate politician Ahmad Chalabi and his Iraqi National Congress, much of which was later found to be false.

Rumsfeld reportedly saw ousting Hussein and establishing an Iraqi government aligned with U.S. interests as the key to changing the entire Middle East region. Moreover, by September 2002, the Bush administration had outlined a new foreign policy strategy, known as the Bush Doctrine, which called for preemptive war to prevent terrorists or state sponsors of terrorism from obtaining weapons of mass destruction. The Bush Doctrine also held that the United States would act unilaterally if necessary to guarantee that the United States remained the sole superpower in the world.

4.3. Neoconservatism, interventionism and the Iraq war of 2003

President Bush began to make the case publicly for military action against Iraq in his January 2002 State of the Union speech in which he identified Iraq as a member of an "axis of evil," along with neighboring Iran and North Korea. All three nations, Bush said, were threatening global security. The Bush administration viewed Iraq as a rogue state and Hussein as a regional troublemaker in the volatile Middle East. Iraq, like many Arab states, opposed Israel, a U.S. ally, and supported the Palestinian cause.

By July 2002 the Bush administration had decided that military action against Iraq was inevitable, according to a British government memo, known as the Downing Street Memo¹² after it was leaked to a British newspaper. Although the Bush administration was publicly proclaiming at the time that war was "a last resort," the memo revealed that the Bush administration had "no patience" for going through the United Nations and that detailed military planning was taking place between the U.S. and British military commanders. The Downing Street Memo stated: "Bush wanted to remove Saddam, through military action, justified by the conjunction of terrorism and WMD [weapons of mass destruction]. But the intelligence and facts were being fixed around the policy."¹³.

In the U.S.-Iraq war the military action begun in 2003 with a United States invasion of Iraq, then ruled by the authoritarian regime of Saddam Hussein. The invasion led to a protracted U.S. occupation of Iraq and the birth of a guerrilla insurgency against the occupation. The resulting destabilization of Iraq also created conditions for a civil war to break out between Iraq's majority Shia Muslim population and its minority Sunni Muslim population. In addition to attempting to quell the insurgency, U.S. forces also found themselves trying to police the civil war. By 2007 the U.S. war in Iraq had lasted longer than U.S. involvement in World War II.

From an abstract point of view, the situation was extremely difficult to handle. President Bush faced two realities. He was not dealing with a nation state that could be defeated by military force, and his attackers could not be deterred by fear of

¹² There were several memos leaked to the press regarding the issue of Iraq and most of them are available here <u>http://downingstreetmemo.com/memos.html</u>;

¹³ Ibidem;

retaliation -- they had to be arrested and incarcerated for an indefinite period, or killed. Even this, however, would not be sufficient. The al Qaeda ideology springs from failed societies and failed cultures; as long as the conditions that produced this cancer continued to exist, it would not be possible to eliminate the threat of further attacks.

Bush chose to use the idea of freedom and democracy -- the American ideology -- as a weapon. Iraq was a target not only because it was a potential source of weapons of mass destruction for the terrorists and a threat to the stability of the region, but because its population was well educated, relatively secular in outlook among the Arabs, and one of the Arab populations most likely to be capable of self-government.

The Bush doctrine commenced its plight by stating that the United States of America are the sole power and they are in need to preserve this privileged position, whatever the costs. The United States have been endowed with excellent military capabilities that need to be preserved in order to be able to respond to all possible challenges¹⁴. In the Bush administration's view the most powerful threats came from rogue states and terrorists armed with weapons of mass destruction.

Here is what the National Security Strategy argues on the matter - "we must be prepared to stop rogue states and their terrorist clients before they are able to threaten or use weapons of mass destruction against the United States and our allies and friends...We must deter and defend against the threat before it is unleashed"¹⁵.

It is the view of neoconservatives such as William Kristol and Robert Kagan that American might (be it political and especially military) is a precondition for worldwide stability. Moreover, the Bush Doctrine's emphasis on the American hegemony/empire binds all the neoconservative aspects together. Neoconservatives utterly dislike the traditional balance-of-power politics and are in favour more of the politics of bandwagoning that puts America in the leading position.

There are several elements that make up the Bush doctrine: the identification of the threat – rogue states and terrorism, followed by the need for a preemptive action to protect the country's integrity. A third element prompts forward a unilateral action. The last element of this interesting doctrine is the promotion of democracy. In this respect, the doctrine might be considered as stemming from Woodrow Wilson's idealism. However, a closer look to the previous characteristics displayed above, one would easily point out the degree of aggression existent inside the doctrine. The best example in this case is that of the Iraq war of 2003, which was intended to be the first step in democratizing the entire Middle East.

On the other side, realists argued that a war was unnecessary since Iraq could be contained in a peaceful manner. Moreover, they emphasized the fact that a war with Iraq for the demise of Saddam and humanitarian reasons would only divert attention from the real threat posed by Al-Qaeda. Realists focus especially on the concept of national interest, which, in the case of the United States, must arise from its needs as a country to ensure the protection of its citizens and territory. Apart from

¹⁴ President Bush's speech at Westpoint, June 2002, available at <u>http://ics.leeds.ac.uk/papers/vp01.cfm?outfit=pmt&folder=339&paper=380</u>

¹⁵ National Security Strategy 2002 - available at http://www.whitehouse.gov/sites/default/files/rss_viewer/national_security_strategy.pdf;

this, national interest has to be proportionate with the power the respective state displays.

What is more, realists contradict the need of building an American empire and militate for a reduction in military interventions worldwide. As far as the policy of bandwagoning is concerned, realists, such as John Mearshimer dismiss it on account of its lack of coherence and the multiple examples of failure; such as Iran and North Korea, countries which utterly refused to take the American side in the latter's involvement in the Middle East¹⁶.

As far as realists are concerned, a preventive war can easily be replaced by a severe containment action, the latter favoured by a weakened Iraq under economic sanctions.

With regard to the matter of interventionism and the Iraq war, things have taken a wrong path. The aim of the Chicago speech was to generate an intense debate about the conditions that need to be met in order to legitimize an intervention. As mentioned previously, the aim of such an intervention was to protect worldwide recognized moral values and not to capitalize on territories. Although Prime-minister Blair attempted to transform Iraq 2003 into a humanitarian intervention this was not possible, since the Iraq war had not been fought on these grounds. Unlike the American counterpart, Tony Blair had to pursue a longer road to ensure British support for such an intervention. And the rhetoric of interventionism was not sufficient enough. At the time Blair had to use the incentives of reopening the Peace Process for the Middle East and an eventual solution for the Israel –Palestine issue to bargain for both the Parliament's and the Cabinet's support.

Critics argue that the doctrine of the international community was to be allotted only to those interventions that had a humanitarian aspect. That is why it was easily to intervene in Sierra Leone in 2000 on the basis of this doctrine, but not that easily to intervene in Afghanistan or Iraq on the same rhetoric.

"This so-called right or duty of humanitarian intervention refers, in such instances, to action taken against a state or its leaders, without their consent, for purposes which are claimed to be humanitarian or protective (the most controversial form of such intervention being military)"¹⁷.

In the meantime, attention shifted from the issue of humanitarian intervention towards the war on terror. Islamic extremism and terrorism are two factors that led to the increase in concern on the part of the British Cabinet as Blair often spoke of a global threat that was clear towards state security and the imperative need of countries to counteract it.

In another speech of April 2009, ten years later, Blair took the terrorist events of 9/11, 7/7, Afghanistan and Iraq and turned them into a case in point for engaging the Muslim world into curtailing Islamic violence. According to the latter, the doctrine still needs to remain in place, as then more than ever there was a need to apply its guidelines. It was the context demanding for the application of the doctrine that changed dramatically.

¹⁶ Michael Williams and Brian C. Schmidt, *The Bush Doctrine and the Iraq War: Neoconservatives vs. Realists*, paper presented by Michael C. Williams at the American University of Paris on Wednesday the 13th of February 2008 pp.16-17;

¹⁷ Chriss Abbott, Rights and Responsibilities: The Dilemma of Humanitarian Intervention, in Global Dialogue, Volume 7, Number 1–2, Winter/Spring 2005—Humanitarian Intervention;

Finally, it must be stated clearly that the Iraq war of 2003 was not a type of humanitarian intervention, although after the failure in locating weapons of mass destruction, there were notable attempts to transform it in one. These kinds of attempts can only destroy the credibility of a humanitarian intervention per se.

5. Conclusions

It is a fact that nowadays, the American support for the war, after the end of the military operations and in light of the need to have a continued presence in Iraq has reached interesting low levels. Why is that possible? Because hoping for democracy in Iraq is pure utopia. It is a mistake that the Bush administration along with the Neoconservatives committed and that needs to repaired by the followers, namely the Obama administration.

The responsibility for pushing forward with the democratization of Iraq and the entire Middle East falls within the ranks of the neocons of the Bush administration.

It is generally acknowledged that political societies are not a homogeneous work and every new theory has to bear that in mind. This was the case with Iraq and might be the case with other countries of the Middle East.

In the defence of the neoconservatives and the American administration, the decision to interfere and stay in Iraq to bring along democracy and self-rule was taken based on the assumption that such an action posed to obvious challenges.

The aim of this paper was to portray the decision of going to war in Iraq throughout the frame of two opposing doctrines. Even if there were attempts to find similarities between the two they still remain different, having different backgrounds, ideas and a whole different set of people to propagate them. Our intensions were to identify the manner in which the Iraq war was represented in the two countries through the prism of these doctrines. Moreover, we have set the goal of analyzing the manner in which the decision-making units used the basic points of these doctrines in order to justify their choice of going to war in Iraq.

The results of our research have pointed out the fact that while the Bush administration relied successfully on the precepts of neoconservatism and borrowed its ideas to make Iraq a case in point, the Blair Cabinet failed in connecting the doctrine of interventionism with the need for going to war in Iraq.

In the case of the United States, Iraq was the first test for neoconservatism after the end of the Cold War. One might argue that Kosovo could have been a test also, but at the time, President Clinton was very much in favour of the containment policy and the neoconservatives found no ears to listen to their pleas. That is why the Iraq war of 2003 is the first time to put the neoconservative doctrine at work, with the support of the administration. And the results have not been the expected ones.

On the other side of the Atlantic, the interventionist doctrine had already been put into practice several times before Iraq, in Kosovo in 1999 and Sierra Leone, in 2000. These two are accurate examples of humanitarian interventions. But when it came to legitimizing the British military presence in Iraq on account of a humanitarian intervention, the British Parliament and public opinion refused to accept such a justification. Furthermore, the absence of a United Nations resolution to legitimize the military intervention in Iraq has struck the interventionist doctrine and its supporters in their endeavours to adapt it to the moment. Taking everything into consideration, while in the neoconservative rhetoric the Iraq war is portrayed as a just war, seeking to spread democracy to the Middle East and striving to maintain the US as the only true power in a world they see as unipolar, the interventionist rhetoric failed to justify any kind of British presence in Iraq in 2003 and remained a doctrine dealing with its one and only purpose – that of providing the legal background for help in the event of humanitarian interventions.

References

Books

- 1. Dorrien, G. (2004), *Neoconservatism and the New Pax Americana*, Routledge, Oxfordshire, Routledge:
- 2. Fukuyama, F.(2006), *America la răscruce. Democrația, puterea și moștenirea neoconservatoare*, ed. ANTET XX PRESS, Bucharest;
- 3. Halper, S. & J. Clarke, (2004) *America Alone. The Neo-Conservatives and the Global Order*, Cambridge University Press, New York, , 2004;
- 4. Ryan, M. (2005), *Neoconservatism and the New American Century*, Palgrave Macmillan, New York;

Articles

- Abbott, C.(Winter/Spring2005), *Rights and Responsibilities: The Dilemma of Humanitarian Intervention*, in *Global Dialogue*, Volume 7, Number 1–2,—Humanitarian Intervention;
- Meyers, H.(ed.)(Spring-Summer2009), Tony Blair, Doctrine of the International Community: Ten Years Later, in Yale Journal of International Affairs, Spring - Summer 2009;
- Dunne, T.(2005, March 1-5), 'Fighting for Values' Atlanticism, Internationalism and the Blair Doctrine, Paper presented to the ISA Conference Hawaii;
- 4. Fairclough, N.(2007), Blair's contribution to elaborating a new 'doctrine of

international community', in Lilie Chouliaraki(ed), *The soft power of war*, Amsterdam, John Benjamins;

- Fukuyama, F. (February 19 ,2006), *After Neoconservatism*, available at http://www.nytimes.com/2006/02/19/magazine/neo.html?pagewanted=all;
- Haines, S. (2010), 'A Word Full Of Terror To The British Mind': The Blair Doctrine And British Defence Policy, in D Brown (Ed), The Development of British Defence Policy: Blair, Brown and Beyond, Ashgate, UK ;
- 7. Kesler, R. C.(Summer 2007), *Iraq and the Neoconservatives*, in *Claremont Review of Books*;
- Dr. Ralph, J., Another doctrine of international community? UK foreign policy and the Obama administration, available at <u>http://www.polis.leeds.ac.uk/assets/files/research/working-papers/ralph-obama-ia-2-Feb-09.pdf;</u>
- 9. Dr. Ralph, J. (August 2005), *Tony Blair's 'new doctrine of international community' and the UK decision to invade Iraq*, in *POLIS Working Paper* No. 20, Leeds;
- 10. Jane E. S. (Jul., 2003) , *Law and Force after Iraq: A Transitional Moment*, in *The American Journal of International Law*, Vol. 97, No. 3, pp. 628-642;
- Peter J. Wallison, R., *Iraq, and Neoconservatism*, available at <u>http://spectator.org/archives/2004/06/16/reagan-iraq-and-neoconservatis/1;</u>
- Wheeler J., N. and T. Dunne (Oct., 2001), *East Timor and the New Humanitarian Interventionism*, in *International Affairs* (Royal Institute of International Affairs 1944-), Vol. 77, No. 4 pp. 805-827;
- Williams C., M. & B. C. Schmidt (February 13, 2008), *The Bush Doctrine* and the Iraq War: Neoconservatives vs. Realists, paper presented by Michael C. Williams at the American University of Paris;

- ***, Tony Blair on Terrorism and 'International Community', in The News, Issue No. 76, March/April 2004;
- ***, Doctrine of the International Community, Chicago Economic Club, April 22nd, 1999, available at <u>http://www.pbs.org/newshour/bb/international/jan-june99/blair_doctrine4-</u>23.html;
- 16. ***, National Security Strategy of the United States of America, January 2002, available at http://www.whitehouse.gov/sites/default/files/rss_viewer/national_security_s_trategy.pdf;
- 17. ***, *President Bush's speech at Westpoint*, June 2002, available at http://ics.leeds.ac.uk/papers/vp01.cfm?outfit=pmt&folder=339&paper=380.

Jewish converts, their communities and experiences of social inclusion and exclusion in post-apartheid South Africa.

Elli Kriel

PhD candidate, Department of Sociology, University of Cape Town, South Africa. Lecturer, Department of Sociology, University of Johannesburg, South Africa. ellik@uj.ac.za

Abstract. Apartheid created rigid boundaries between ethnic and racial groups in South Africa and restricted the movement of people between groups. With regard to conversion into the Jewish community in South Africa (a white English-speaking community), this meant that a contracted cohort of white individuals only were permitted entry into the community. However, the collapse of apartheid in 1994 marks a change in the conversion trends to Judaism in South Africa and in the last especially, black and Afrikaans converts decade have increasingly sought conversion to Judaism. However, it is reported that the conversion process has become stricter at the same time as researchers claim that the community is becoming more insular. This study will investigate the reasons for the conversion changes and will question to what extent converts feel included or excluded in the community. The study will furthermore explore if conversion in post-apartheid South Africa is an expression of identity politics in the 'new South Africa' as it was during apartheid. The methodology will be social constructionist using a narrative approach and qualitative interviews with converts and rabbis will be used to collect data. Discourse, practices, agency, and subjectivity will be explored through the narratives provided by converts and rabbis. The study, which is still at the conceptualisation stage, is important as it will be the first to give a voice to converts in the community. It will also reflect on the multicultural nature of South Africa's democracy as it pertains to this ethnic and religious minority group.

Keywords: conversion, Judaism, South Africa, apartheid, postapartheid.

1 Introduction

South Africa is a place where race and ethnicity have had, and continue to have, important consequences for social exclusion, inclusion and citizenship of its inhabitants. Since the birth of the modern South African state, as with all communities within its borders, race relations have impacted on the formation and establishment of the Jewish community in South Africa (Beinart, 1996; Herman, 2007; Kosmin, Goldberg, Shain & Bruk, 1999; Krut, 1987; Marks, 2004; Mendelsohn & Shain, 2008; Shain, 2011; Shimoni, 1980). The Jewish community emerged from a diverse immigrant base deeply invested in creating and maintaining a uniquely blended but homogenous ethnic character and religious identity for all its members. The dominant identity of SA Jewry materialised into an amalgamation of Anglo culture with Eastern European religious values and political ideologies. The form it took is that of a white, urban, middle-class community that is English-speaking, of Jewish Orthodox denomination, highly Zionistic, and with communist leanings at times. This sociocultural formation, which is embedded within the contemporary community, developed not only out of various cultural backgrounds of South African Jewry, but also as a social and political reaction to the unfolding race and class dynamics within the State over the last century.

In South Africa, the issue of conversion has arisen for the Jewish community as it has attempted to strengthen its social, political and religious interests and boundaries (Zekry 1998). This was most acute during apartheid, by which time the community's identity had already formed in response to the conflict between colonial and Boer powers. Indeed, the ideology of apartheid further supported the separate and homogeneic development of communities and apartheid law created rigid boundaries between different ethnic groups. It is unsurprising, therefore, that the community was invested in maintaining a 'homogenous' white identity and candidates for conversion were thus narrowly selected.

Conversion was never encouraged in the first place (Zekry, 1998), but apartheid simplified the issue for the community. Community prohibitions on conversion existed at various times on the basis of race and reason (marriage was not considered adequate; religious conviction was (Zekry, 1998)). As a result a contracted cohort of only white candidates were permitted entry into the community via conversion (Zekry, 1998). The conversion of black individuals was especially difficult - if not impossible - given the legal restrictions on race groups, i.e., in terms of their interpersonal relationships, movement, where they could live, pray, congregate, and attend school. Requests for conversion during apartheid arose mostly for marriage (even though discouraged, if not prohibited) and less for religious conviction (possibly because of the strong Christian indoctrination of the South African state). Conversion was very selective; inclusion was conditional on specific social demands even in the absence of conversion (such as raising children as Jewish, adopting a Jewish lifestyle etc) but formal acceptance in the community was nonetheless uncertain if not entirely lacking at times (Zekry, 1998).

The transition to democracy in 1994 was a pivotal moment in South Africa's history, and it impacted on the political outlook of the Jewish community in maintaining a place for itself in the 'new South Africa'. Jewish community researchers assert that, despite the collapse of apartheid and almost two decades of

democracy, the community is becoming more insular, inward looking, religiously conservative and that identity adaptation to a new social environment has not occurred (Herman 2007; Shain, 2011). At the same time, democracy and the repeal of legislation compelling closed communities has introduced a significant shift in conversion trends which reflects the multicultural nature of contemporary social relations in the country. Since the early 2000s, converts from all race and language groups are presenting (Hendler, 2011), primarily for religious conviction and, in some cases, already as family units (Hendler, 2009). The process of conversion is also acknowledged as being stricter, more inflexible, formal, bureaucratic and more procedural than in the past. Apartheid contributed to the cohesiveness of the community, a homogenous identity and stiff boundaries: conversion occurred within this framework. Democracy under black majority rule is new and unfamiliar territory for white minorities. Is this why 'new' convert types are presenting and the conversion process has become tighter and more formalised? How has the Jewish community responded to conversion in a post-apartheid environment - now that the environment is fundamentally different? In this context, an exceptional opportunity exists to understand the transformation of the social identity of converts, the agency of converts to negotiate new norms within seemingly rigid social and cultural categories, ethnic cultural formation through conversion processes, and community adaptation in the post-apartheid environment.

Overall this study investigates the process of converts' identity transformation and questions their experiences of social inclusion and exclusion their 'citizenship' - in the current Jewish community. If conversion to Judaism is in essence more than a religious experience, and rather a social process of how the community deals with 'outsiders', regulating whether (and which) new entrants may be received, then it is apt to research how converts may come to 'fit' into the established self-ascribed homogenous character of the community particularly when the community no longer defines itself in racial terms. In the case of the converts, the central question this study asks is: what is required of a convert to achieve approval and gain entry into the community? It further probes when formal acceptance by religious authorities is achieved and communal participation approved of. In this regard, the voice of converts is critical. In the case of the community, is the conversion process about maintaining a homogenous tight-knit community within a changed political environment? Under what circumstances is cultural change permissible, what conditions allow for communal adaptation, and what are the drivers for this adaptation? On this point, the study investigates how the religious authorities of the Ecclesiastical Court (Beth Din) respond to 'atypical converts' of various race, ethnic, class and language backgrounds in post-apartheid South Africa. The research also asks, to what extent are 'atypical converts' socially acceptable in the postapartheid Jewish community? How has the process adapted? The study turns to converts in the only two locations where conversions are permitted in South Africa -Johannesburg and Cape Town - as well as conversion authorities in both areas, for the period 1994-2012, in order to answer the following research questions:

1. How do converts to Judaism experience social inclusion and exclusion in the Johannesburg and Cape Town Jewish communities in post-apartheid South Africa?

This question will be conceptualised as the 'first level' of the study on social inclusion and exclusion as it relates to the micro level. It focuses on those who cross an imaginary boundary of who may be a member of the community and explores the adoption of a perceived identity as a 'typical South African Jew'. The experiences and agency associated with a new way of being will be interpreted. The relationship between the diversity of candidates (in terms of race, class, cultural background, language, sexuality and marital status) and the extent to which social inclusion is experienced will also be raised.

2. How do the Johannesburg and Cape Town Jewish authorities' manage conversion in post-apartheid South Africa? What are their views of adaptation of their perceived ethnic identity to the broader post-apartheid social environment?

The above research question will be conceptualised as the 'second level' of social inclusion as it relates to a broader level of conversion - the communal level. It questions how the movements around religious conversion can be understood with reference to the community's experience and perceptions of their social inclusion in the 'new South Africa'. The question probes the community's insecurities in post-apartheid South Africa and the relative contribution conversion makes to the community in this regard.

It must be noted that the South African Jewish community is divided into two denominations namely Orthodox and Reform (also known as Progressive). Although the Orthodox community dominates, there are almost as many converts to Reform Judaism as there are to Orthodox Judaism. Therefore the terms 'converts' and 'conversion authorities' refer to those of both denominations. Furthermore, there are many reasons for conversion including religious conviction, marriage, Jewish history outside of matrilineal descent which 'disqualifies' one from Orthodox Judaism, and adoption (Zekry, 1998). This study will focus only on those converts for religious conviction and marriage as it assumes less exposure to Jewish culture, networks and identity. Based on the above research questions, the hypotheses will question:

Do converts to Orthodox Judaism in Johannesburg and Cape Town feel more socially included than converts to Reform Judaism in Johannesburg and Cape Town?

Do both the Orthodox and Reform conversion processes in Johannesburg and Cape Town facilitate only marginal social inclusion in the community?

Is there a relationship between race and first language of converts and their experiences of social inclusion or exclusion during and after the conversion process to Orthodox and Reform Judaism in Johannesburg and Cape Town?

On what social factors is the social inclusion of converts to Orthodox and Reform Judaism conditional e.g. age, marital status, choice of marital partner, choice of synagogue and congregation, level of religious observance, receiving family's level of observance (if applicable), and nature of children's upbringing? What broader socio-political factors influence the views of conversion of the Orthodox and Reform religious authorities in Johannesburg and Cape Town?

Do the Orthodox and Reform religious authorities of conversion in Johannesburg and Cape Town encourage or discourage diversity of conversion candidates into the community and why?

2 Literature review

Religious conversion in general is a topic that has received enormous scholarly attention in a number of disciplines including biblical studies, theology, sociology, anthropology and psychology. In terms of sociology, the discipline of concern here, the interest around conversion has been primarily in explaining models of conversion with regard to the acquisition and retention of members in the rise of new religious movements and cults within modern and democratic systems. Religious conversion to Judaism is naturally of specific interest to Judaic studies both in Israel and the diaspora, as it relates to religious-legal definitions of group membership and identity i.e. who/what is a Jew? While the content of the debates around conversion in Israel and the diaspora are undoubtedly different (given the particular contextual factors that weigh on the issue), the concerns around citizenship within a community based on valid membership and identity are the same. Conversion to Judaism therefore has been an issue that has been much studied by Jewish scholars around the world.

Connecting Jewish conversion to identity essentially implies that there is a concern about group boundaries and membership in order to maintain a lasting ethnic formation. However, group membership is challenged by intermarriage and assimilation and therefore conversion becomes a mechanism for retaining members and ethnic boundaries. In the United States, for example, where the largest Jewish population in the world is found outside of Israel, the issue of conversion has received enormous attention from the Orthodox, Conservative and Reform Jewish movements because of the concern over assimilation and intermarriage and the disappearance of the American Jew. In the South African Jewish community however there has been very little assimilation of Jews, high levels of endogamy and low levels of intermarriage. Hence, the conversion numbers are small¹ but so too the community is small. Conversion as an issue has always been treated with importance but has not been very prominent within the community. This is reflected in the lack of studies on the topic of conversion in South Africa.

It appears that only two major studies relating to conversion have been conducted in South Africa. The first, an historiographic study titled *The Nature and Development of Orthodox Judaism in South Africa*, circa 1933 (Simon, 1996), contains a chapter dedicated to conversion. The chapter examines the topic from its earliest beginning in South Africa but ends in 1933. The second study is a 1998 PhD study *Conversion in South Africa by* Rabbi Pinchas Zekry. It is arguably the most

¹ It is estimated that the total number of converts in the community (irrespective of when conversion took place) is 4% (Kaplan, 2006; Kosmin et al, 1999)

comprehensive contribution to the topic of Orthodox conversion in South Africa, also covering the early *Beth Din* (Ecclesiastical Court) records, but focusing quantitatively on the issue mainly from the 1950s through to the mid-1990s. While the study is important in understanding the establishment of the Orthodox conversion process historically and contextually in South Africa, and provides the reasons for conversion, the characteristics of the converts, and the views of some converts in the Durban Jewish community towards conversion (the Durban community is the community of the author of the study, of which he is the Rabbi), it is concluded in 1998, more than a decade ago. To my knowledge, while numerous studies have been conducted on the South African Jewish community, no further study has since been conducted on Orthodox conversion.

Both studies show that in South Africa conversion is as much about "who is a Jew" as it is about the "Jewish Question", namely the question of the citizenship rights of Jews in emerging modern states with the rise of the western democratic systems. Though focused on different time periods, the two studies show that conversion has reflected the issues that the community was contending with as the South African state developed and modernised. Indeed, according to Zekry (1998) the conversion trends of the last century have been very sensitive to the 'mood' of the community. Authorities throughout the community's development were clearly influenced by anxiety relating to acceptance in South Africa and the same could be expected of today's situation. The views of conversion seem to have been an extension of the feelings of insecurity of the community having initially come from various European origins trying to settle and adapt to a new (safe) environment at the same time that the South African state was emerging. The situation is neatly summarised by Simon (1996:135) who argues that '[b]y and large, it seems that the Jewish attitude towards proselytism has always been inversely proportionate to prevailing feelings of comfort of security and self-confidence. When the community felt itself under threat spiritually or physically it closed ranks and resisted the would-be proselyte. A community sure of itself and its strength (even if mistakenly so) was more likely to be willing to receive new adherents'.

While both Simon's (1996) and Zekry's (1998) studies reflect on the early post-apartheid period, which is when they were published, racial issues are neglected possibly because the trends that have emerged were not yet showing in the conversion process and also because it was not the focus of their studies. An article in a recent edition of the community journal *Jewish Affairs* (2011), indicates that '(f)rom Ghana to Krugersdorp, from Newcastle to Ethiopia, Israelis, Indians, English, Afrikaans or Sotho speaking singles, and entire families, there seems to be an increasing demand for conversions to [Orthodox] Judaism in South Africa' (Hendler, 2011:30). It is time therefore that the topic be reexamined. Why has interest in Orthodox Judaism supposedly increased in the abovementioned ways? Why especially given that empirical and anecdotal evidence suggests that the Orthodox process of conversion has become stricter in the decades of democracy (which has introduced reconciliation, diversity, multi-culturalism and non-racialism)?

At the same time, it is important to note, that according to the 2001 South African Population Census there 11979 members of the black population in South Africa who claim to be Jewish (Statistics SA, Census, 2001) but who do not appear in 'mainstream' Jewish circles/ life. Since it is understood that they have not converted

(if we accept the statistics of Beth Din as presented by Zekry (1998)) it highlights that there are racially discrete understandings of what it is to be a Jew in South Africa. What happens when this racial divide is bridged through conversion? What do converts of 'colour' - and other 'non-traditional' converts for that matter - experience in the community?

There are also a number of other gaps that are apparent. Firstly, both studies exclude the Reform community. In the case of Simon's (1996) study the Reform movement was not yet a relevant issue as it was only formed in 1933. According to Zekry (1998), attempts were made to include the Reform community in his study however he was unsuccessful in being able to do so. It appears that there are no studies on Reform conversion in South Africa although there may be some material in existence within the Reform community that has not been made public. The lack of research on Reform conversions may be attributed to a number of reasons including the marginality of the Reform community, the newly centralised nature of its conversion process or the unwillingness to share information with Orthodox authorities who have publicly scorned them.

Secondly, the studies that exist offer a limited view of conversion from the perspective of converts. That which does exist is of a quantitative nature and is drawn from a closed-ended questionnaire. It is also based on a few converts in the Durban community only which skews the data as the Johannesburg and Cape Town dynamics are different (by Zekry's own admission) and the converts are not exposed to the same level of stringencies as converts in other regions.

More than a decade has lapsed since the aforementioned studies. In the time that has past, the trends in conversion seem to be changing (Hendler, 2011) and a number of gaps exist in what is known about conversion to Judaism in South Africa. Therefore, despite the few previous studies, an updated sociological analysis is required in order to understand conversion as a community process and whether/how the post-apartheid environment is changing the kinds of applicants to the process, the nature of conversion, the converts' experiences of social inclusion and conversion as an expression of the community's participation in South Africa's democracy. It is also important because it will be the first to give a more inclusive and in-depth voice to converts. The assumption here is that the narratives of converts are critical in coming to see how identity transformation through conversion occurs. The recognition exists that a convert's belonging to any particular type (e.g. race or language group) will alter the voice of the convert and the construction of the narrative.

2.2. Research problems and objectives: key questions to be asked

The research questions are formulated as follows:

- 1. In post-apartheid South Africa, since 1994 until 2012, who is converting to Orthodox and Reform Judaism and why? In the decades of democracy, what conversion patterns are presenting compared to the community's current demographic trends and political purposes? How do the statistics differ to pre-1994 statistics?
- 2. What are the experiences of social inclusion and exclusion of Orthodox and Reform converts in the post-apartheid Jewish community and what are the bases for similarities and differences in the experiences of various converts?

Does race, language, gender, former religious beliefs, marital status, conversion with a 'ready-made' family (partner and children), and age affect experiences of social inclusion and exclusion? Do Afrikaans converts have better experiences than other converts? Do black converts have more negative experiences than other converts?

- 3. Does the entry of various kinds of converts particularly Afrikaans and black converts contribute to diversity in the community? Is cultural, racial, linguistic and sexual diversity permitted and/or tolerated in the community, when, and in what form? Do converts' identities adapt to the ethnic homogeneity of the community or in some way 'change' it? What 'outside' influences are permitted into the community through conversion?
- 4. In post-apartheid South Africa, what are the distinct approaches of the Orthodox and Reform Jewish authorities in respect of conversion? And, what impact do these differences have on the experience and perception of social inclusion and exclusion of converts in Johannesburg and Cape Town?
- 5. How have Orthodox and Reform conversion authorities responded to the dynamics of the community in the post-apartheid environment? Does conversion strengthen its boundaries and content? If so, how? If not, why? Since the numbers of converts into the community are small it is hardly expected to make a quantitative difference to the community. Therefore in terms of its qualitative nature, what does conversion contribute to the community?

2.3. Research problems and objectives: broader issues to be investigated

The study will first and foremost address the voice of converts. A study on conversion in South Africa, in which the voice is revealed of a diverse range of converts into a small and generally rigidly defined community, will make a rich research project. It will be the first to provide a platform for many different kinds of converts to express their perceptions of their social inclusion or exclusion and stories of their experiences within the community as converts. Anecdotal evidence suggests, and personal experience as a convert into the community finds, the voice of converts is muted (for some more than others) and there is a great need to address the issue. The results will therefore be important for converts, rabbis and their communities in facilitating a more inclusive approach to conversion. A study of this kind will also contribute to literature on the topic of Jewish conversion in South Africa which is lacking at a scholarly level.

The study furthermore addresses the value of conversion at a communal level and how this may be understood within a 'new' South African context. The focus is directed at broader South African socio-political factors and how they may inadvertently influence community membership 'projects'. In particular, this study will make a contribution to observing how the turn to democracy and reconciliation in post-apartheid South Africa has impacted on various religions within the state (resulting in the decline of some religious institutions such as the Dutch Reformed Church and the growth or intensification of others, such as Judaism). Within this framework the manner in which faith communities – such as the Jewish community - have responded to a completely different socio-political environment will be important.

It is quite unusual for a study on conversion to focus on how it is part of a communal process of adaptation to the broader social environment in which the community is found. In doing so, the research questions also relate to the broader problem that has preoccupied Jews since enlightenment, and which is sensed in South Africa today, of 'how to enjoy full citizenship without losing their separate identity' (Beinart, 1996:65). This question has received considerable attention from many great classical and philosophical thinkers including among many others Karl Marx, Max Weber, Immanuel Kant, Hannah Arendt, and Walter Benjamin. It is, furthermore, a lens to examine how ethnic minorities adapt to their external pressures in plural societies and can be applied beyond the Jewish community to other minority groups. It also relates to broader issues of democracy and cosmopolitanism – how do minority groups negotiate their insecurities in democracies, what responses might they consider in securing their inclusion and what does the South African example show us?

2.4. Principle theoretical frameworks (in brief)

Conversion, as an expression of a community adapting to a modernising state, will be situated within the classical modernisation theories of Marx, Durkheim, Weber. Furthermore, Talcott Parsons' structural functional analysis of modern society and pattern-variables, as well as Eisenstadt's (1966) views of modernisation and how it is that traditional societies transition and adapt, will also be relevant. In this regard, Ferdinand Toennies' distinction between gemeinschaft and gesellschaft may also be helpful. So may Frederick Barth's (1969) work on ethnic boundaries in which exclusion devices are examined. Other perspectives on situational ethnicity (Gluckman, 1958 cited in MacMillan, 2000) will also be explored. The theories of Ulrich Beck, Anthony Giddens and Zygmunt Bauman will be referred to for more pluralistic understandings of modernity. Benedict Andersons (1991) Imagined *Communities* in which the cultural roots of nationalism are theorised will also be reflected upon. Gloria Anzaldua's (1987) theory of Borderlands which looks at multiple, socially constructed borders (such as physical, psychological, spiritual and sexual borders) between individuals and groups will also be delved into. The concept of liminality (Turner, 1969) will be useful in understanding the converts' experiences of conversion.

The study will also enter current debates on secularisation theories within the sociology of religion (Kaufman, 2010). Within this, a number of sociological, anthropological and theological studies have been conducted internationally on conversion and models of conversion (Austin-Broos, 2003; Buckser, & Glazier, 2003; Gooren, 2007; Rambo, 1993; Snow & Machalek, 1984; Sremac, 2010). Many of these studies focus on conversion to new and proselytising religious movements – both of which Judaism is not – but are nonetheless useful in understanding conversion as a communal process. In terms of thinking about the citizenship of Jews in the diaspora this thesis will be framed within debates around citizenship, modernity, democracy and cosmopolitanism. In linking the study to contemporary South Africa, Ari Sita's (2010) work on the *ethic of reconciliation* and *symbolic figuration* will be included in relating this project to the post-apartheid era. Furthermore, the changes in the

religious landscape of South Africa in the post-apartheid will be made sense of in terms of the various works of Jean and Joan Comaroff.

3 Data and methodology

The methodological framework for this study will be social constructionist. Social constructionism is a fairly new methodological application to the study of religious conversions (Sremac, 2010). Previous studies have tended to rely on psycho-theological approaches which promote 'internal psychological processes of the individual's conversion experience' and have tended to avoid identity and experience (Sremac, 2010:8). Since identity and experiences are at the core of this study, social constructionism is a useful framework especially since it permits a 'more contextual-communal understanding of the conversion phenomena' (Sremac, 2010:8) by placing emphasis on narratives which 'describe and interpret the processes of conversion phenomena' (Sremac, 2010:8).

However, the research design will include mixed methods. Firstly, it will make use of existing quantitative data from secondary sources in order to paint a picture of conversion to Judaism within South Africa. The data sets that will be accessed are Rabbi Zekry's (1998) dissertation which is a collation of all the Ecclesiastical Court (Beth Din) records for the Orthodox community. This will be followed with a request for updated statistics of conversions since 1998 until present from the Orthodox Beth Din. The updated Orthodox statistics will help clarify if and what new patterns of conversion are presenting. In order to get a full sense of conversion to Judaism in South Africa statistics from the Reform community will also be requested. However, as the centralisation of the Reform community's conversion process through the Beth Din is still fairly new, it is not known whether or how much data exists on Reform conversions at the current time. Other data that will be accessed is the 2006 attitudinal survey of the identity of the Jewish community in postapartheid South Africa conducted by Kaplan Centre for Jewish Studies and Research (UCT). Although the usage of quantitative data within constructionist methodological frameworks is unusual, the usage of these statistics will assist in deepening knowledge of the historical context and culture-specific detail for conversion, especially in the post-apartheid environment. This quantitative data is essential for establishing part of the contextual-communal understanding of conversion as part of the social constructionist approach.

The second research method will be qualitative and this aspect of the study will make up the bulk of the analysis and interpretation. It will capture the narratives of converts and rabbinical authorities through interviews. The assumption is made that the 'self' is dependent on social context and identity is formed through social interaction. Based on this assumption, it is further understood that the social processes in a reference group are the cultural 'vehicles' for the making (transformation or reconstruction) of social identity (Sremac, 2010). Discourse, as exposed in narratives, is therefore important as it represents expressions of identity and the self, and meaning is ascribed to these experiences which help to form the self. In order to understand how identity is formed and experiences through conversion are

understood, narratives are important to this study. It will work well at a micro level, but also at a communal level with understanding the symbolic figuration (i.e. identity reconfiguration) of the Jewish community in post-apartheid South Africa. A narrative approach to the data collection and analysis therefore lends itself well to the research questions, methodological framework and the qualitative research method chosen. Researchers are advocating for its greater use in conversion research as it stresses discourse which it locates in historical and cultural context (Sremac, 2010).

The specific qualitative method to capture narratives will be in-depth, semistructured interviews with participants from the Orthodox and Reform communities in Johannesburg and Cape Town. A pilot study will be conducted with six converts in order to help establish the best possible interview schedule. Thereafter, the participants will be converts who have lived as Jews in South Africa for at least two years, rabbis responsible for the conversion process (e.g. registrars of conversion), rabbis of the two *Beth Dins* who permit final conversions, and also community rabbis who facilitate the entry of converts into actual communities. The relevant rabbis will be purposively identified through the *Beth Dins*. Purposive sampling will also be employed to identify and access one convert. Thereafter, snowball sampling will be used to access other converts. Converts and rabbis will be sought accordingly:

Converts	Orth	nodox	Re	eform
	JHB	СРТ	JHB	СРТ
Race (black including	5	5	5	5
Indian/Asian, coloured)				
Language (Afrikaans)	5	5	5	5
Reason for conversion marriage	5	5	5	5
Sub-total	15	15	15	15
Rabbis	Orthodox		Reform	
	JHB	СРТ	JHB	СРТ
Registrar	JHB 1	CPT 1	JHB 1	CPT 1
Registrar Judges of the Beth Din	JHB 1 3	CPT 1 3	JHB 1 3	CPT 1 3
5	1	CPT 1 3 3	JHB 1 3 3	1
Judges of the Beth Din	1 3	1 3	1 3	1 3

Table 1: Table showing the sample composition according to participant type.

A set of semi-structured, probing questions will be developed for each research question, namely experiences of social inclusion and exclusion (referred to above as the first level of analysis) and authority responses to conversion (referred to as the second level of analysis). The interview questions will keep in mind the fact that acceptance after conversion in the past has been conditional on behaviours and demonstration of an embodied sense of being a 'good Jew' (raising children as Jewish, living a 'Jewish life' etc). Therefore the line of questioning will focus on meanings of why a convert believes he/she will potentially make a 'good South

African Jew', what makes a convert a 'good South African Jew', and once entrance into the community has been formally achieved, how it can be shown that a convert is a 'a good South African Jew'. Within these categories, the line of questioning will be divided into four broad areas namely *practice* (what converts need to do to be accepted), *agency* (how they adapt and conform to the demands placed on them), *subjectivity* (how they experience their identities, inclusion and exclusion) and *discourse* (how meanings of the self and community are generated, sustained and/or disrupted through conversion; justifications and narratives of conversion) in order to understand identity transformation during the process and after. The same structure will be used for all converts and rabbis, although the line of questioning will differ in various places at different points in the process and also between converts and rabbis. The table below provides examples of the question frameworks for each broad area of the interview schedule. The shaded areas in the tables denote first level analysis and

	Converts: Orthodox (n=30) and Reform (n=30) JHB and CPT			
	Practice	Agency	Subject- ivity	Discourse
Perceptions	What needs to	How do you	Experiences	How is
of	be done to	deal with	of inclusion	participatio
conversion	achieve entry	difficulties	and	n in the
during	into the	in the	exclusion;	process
process &	process, make	process -	where, when,	justified?
after the	progress &	strictness;	from whom,	What
process:	enter the	values; lack	why and how	debates of
why will	community?	of	are these	conversion
you be a	How does this	knowledge;	experiences	are
good Jew	differ	feelings of	understood?	encountered
(process)?	according to	inadequacy;	What	regarding
Why are	the	survival;	would make a	the position
you a good	characteristics	manage	convert in the	of converts
Jew (post-	of the	family	community	in the
conversion)	convert?	relations;	more	community?
?	On what	change in	comfortable?	What
Now I	practices does	networks;	What change	discourse
am a good	ongoing	orientation	is required	is/was
Jew	inclusion,	etc	and why?	encountered
(when)?	exclusion in	How has	Race,	around an
	the	agency	language,	individual's
	community	changed	age,	particular
	rest? How do	over time	sexuality,	situation?

Table 2: showing question frameworks for converts.

the non-shaded the second level of analysis.

	the characteristics	compared to during the	family situation and	What is the language of			
	of the converts challenge or	process?	marital status	belonging?			
	enhance the		explored				
	experiences of		here.				
	inclusion and exclusion? Do						
	practices						
	compensate						
	for lack of						
	inclusion?						
	At what						
	point are practices						
	irrelevant?						
	When is one						
	full recognised						
	as a 'good Jew'?						
Situational		attitudes, expres	sions ever change	e depending			
Variability		· •	presence of partic				
		ersion, commun	ity, presence of fa	amily etc?			
Commun-	Why?			ationities and			
ity in post-	Converts' experiences of practices, agency, subjectivities and discourses around conversion in response to the broader						
apartheid	environment will be explored here. How did the practices						
South	change (narratives), when (historical, contextual and political						
Africa	forces) and why (discourse)? What reasons were provided by the rabbis and the community (discourse)? How were the						
			and explained (su				
			verts, the commun				
	broader socio-po						

Table 3: showing question frameworks for rabbis.

	Orthodox (n=1	Rabbis: 4) and Reform 8 and CPT	(n=14)
Practice	Agency	Subject- ivity	Discourse

Perceptions	What are the	How have	In what	What are the		
of	criteria for	individual		rabbinic.		
01		rabbis	ways is			
conversion	acceptance	140010	conversion	political,		
during	into the	personally	to Judaism	sociological		
process &	process and	promoted or	different to	(identity)		
after the	the	prevented	other forms	community		
process:	community?	converts from	of religious	and social		
why will you	Explain the	entering the	conversion?	discourses		
be a good	origin of the	community?		relevant to		
Jew	choice of	Why?		conversion		
(process)?	criteria and	What agency		in the South		
Why are	the concerns	is permitted		African		
you a good	they address.	of converts in		community?		
Jew (post-		terms of		How do		
conversion)		customs,		Israeli		
?		language,		movements /		
Now I am a		sexual		changes in		
good Jew		practices,		conversion		
(when)?		cultural		affect South		
		expressions,		African		
		children etc		responses to		
				conversion?		
Situational		policies of conv				
Variability		mitted for conver				
	and for what reasons? How is non-sanctioned variability dealt					
	with? What are the actions, remedies and outcomes of non-					
	sanctioned variability? What are the vested interests of the					
	rabbinate? Power as an issue to be explored here.					
Commun-	As per conv	verts' line of quest	tioning on this to	opic.		
ity in post-						
apartheid						
South						
Africa						

4 Conclusions

Generally, the topic of conversion to Judaism arises in the context of Jewish concerns of assimilation and intermarriage, and therefore conversion is considered as important in asserting or protecting Jewish identity and boundaries. This has undeniably been the case in South Africa. However, what is also evident in the South African scenario, is that conversion taps into a central question of how Jews as a distinct ethnic and religious group may achieve/maintain their identity (by avoiding assimilation and intermarriage) while simultaneously enjoying full citizenship rights in a broader and dynamic pluralistic society. Furthermore, it appears that the conversion trends before and after 1994 are different, and have arisen as a result of the interaction between the characteristics of the South African Jewish community and the changing racially defined socio-political environment of South Africa. To what extent may the changes in conversion trends be explained in terms of broader change in South Africa's sociopolitical environment? And, since converts are expected to embody the identity and practices of an entrenched white community, is racial and linguistic diversity in terms of identity encouraged through the conversion process and what are the ramifications for social inclusion and exclusion in the community? This proposal outlines the framework from which the issue of conversion to Judaism in South Africa is explored with regard to social inclusion and exclusion of converts, and the community, within the broader democratic environment of the post-apartheid era. It will be the first study to provide a 'voice' to converts in the community. It will also be the first to provide a view of conversion to Judaism in the post-apartheid environment in South Africa.

References

Austin-Broos, D. 2003. "The Anthropology of Conversion: An Introduction" in Buckser, A. & Glazier, S.D.(eds) *The Anthropology of Religious Conversion*. Rowman & Littlefield, Maryland.

Bath, F. (1969) Ethnic groups and boundaries. Allen & Unwin: London.

Beinart, P. (1996) "The Jews of South Africa". Transition, Vol 71, pages 60-79.

Braude, C. (2001) *Contemporary Jewish writing in South Africa: an anthology*. University of Nebraska Press: Lincoln and London, pages ix-xxv.

Buckser, A. 2003. 'Social Conversion and Group Definition in Jewish Copenhagen' in Buckser, A. & Glazier, S.D. (eds) The Anthropology of Religious Conversion. Rowman & Littlefield.

Comaroff, J. & Comaroff, J.L. (2000) "Millennial Capitalism: First Thoughts on a Second Coming". *Public Culture*, Vol 12, No 2, pages 291-343.

Comaroff, J. & Comaroff, J.L. (2004) "The struggle between the constitution and "Things African". *The Wiser Review*, Vol 1, No 1, page 6.

Eisenstadt, S.N. (1966) *Modernization: Protest and Change*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Giddens, A. (2006) Sociology (5th edition). Polity Press, Cambridge

Gooren, H. (2007) "Reassessing Conventional Approaches to Conversion: Towards a New Synthesis". *Journal for the Scientific Study of Religion*, Vol 46, No 3, pages 337-353.

Hendler, R. (2011) "Demand for conversions to Judaism Soars", *Jewish Tradition*, Rosh Hashanah 201/5772, Publication of the Union of Orthodox Synagogues SA, Orange Grove.

Herman, C. (2007) "The Jewish community in the post-apartheid era: same narrative, different meaning". *Transformation: Critical Perspectives on Southern Africa*, Vol 63, pages 23-44.

Hendler, R. (2009) *Interview with Registrar of Conversions*. Held at the Johannesburg Beth Din: South Africa.

Kaplan, D.E. (1995) "Your People, My People. Conversion to Judaism and Jews by Choice". *Jewish Affairs*, Vol Spring, pages 87-89.

Kaplan, M. (2006) *The Future of South African Jewry: Comments and Observations*. Kaplan Centre for Jewish Studies and Research, Cape Town.

Kaufmann, E. (2010) Shall the Religious Inherit The Earth? Profile Books Limited, London.

Kosmin, B.A., Goldberg, J., Shain, M., & Bruk, S. (1999) "Jews of the 'new South Africa': highlights of the 1998 national survey of South African Jews". Report for the *Institute for Jewish Policy Research* in association with the Kaplan Centre for Jewish Research and Studies, Cape Town.

Krut, R. (1987) "The making of the South African Jewish community in Johannesburg, 1886-1914" in Bozzolli, B. (ed), *Class, Community and Conflict: South African Perspectives.* Ravan Press, Johannesburg.

MacMillan, H. (2000) "From Race to Ethnic Identity: South Central Africa, Social Anthropology and the Shadow of the Holocaust". *Social Dynamics*, Vol 26, No 2, pages 87-115.

Marks, S. (2004) "Apartheid and the Jewish Question: Review Article." Journal of Southern African Studies, Vol 30, No 4, pages 889-900.

Mendelsohn, R. & Shain, M. (2008) *The Jews in South Africa*. Jonathan Ball Publishers, Johannesburg and Cape Town.

Pollack, J. (2009) *The Kasrils Affair: Jews and Minority Politics in Post-Apartheid South Africa*. Juta & Company Limited, Cape Town.

Posel, D. (2011) "SA Jews must face up to complicity". *Mail and Guardian*, 11 November 2011.

Rambo, L.R. (1993). *Understanding Religious Conversion*. Yale University Press, New Haven and London.

Shain, M. (1994) *The Roots of Anti-Semitism in SA*. Charlottesville:University of Virginia Press.

Shain, M. (2011) "Jewish cultures, identities and contingencies: reflections from the South African experience". *European Review of History*, Vol 18, No 1, pages 89-100.

Simon, J. (2002) "Proselytism in the South African Jewish Community: the problem, the responses, the reception". *Jewish Affairs*, Vol Autumn, pages 12-15.

Sitas, A. (2010) The Mandela Decade, 1990-2000: labour, culture and society in post-apartheid South Africa. Unisa Press, Pretoria.

Snow, D.A. & Machalek, R. (1984) "The Sociology of Conversion". *Annual Review of Sociology*, Vol 10, pages 167-190.

Sremac, S. (2010) "Converting into a New Reality: Social Constructionism, Practical Theology and Conversion". *Nova prisutnost*, Vol 8, No 1, pages 7-27.

Statistics South Africa (2001) Primary tables South Africa: Census '96 and 2001compared.Accessed30April2012fromhttp://www.statssa.gov.za/census01/html/RSAPrimary.pdf

Toennies, F. (1971) On Sociology, Pure, Applied and Empirical. University of Chicago Press, Chicago and London.

Turner, V. (1969). *The ritual process: Structure and anti-structure*. Cornell University Press, Ithaca New York.

Zekry, P. (1998) "Conversion to Judaism in South Africa". Unpublished PhD dissertation, Department of Religious Studies, University of South Africa, Pretoria.

Zekry, P. (2000) "Conversion to Judaism in South Africa". *Jewish Affairs*, Vol Spring, pages 27-32.

Neobohemia: a Dispute over New Artistic Spaces in a Postmodern City

Weronika Maćków

Cultural Studies Department, Adam Mickiewicz University, Poznań, Poland weronika_mackow@poczta.fm

Abstract. The following paper introduces the issue of neobohemia and the role of artistic districts in the production of symbolic capital of Polish, as well as Eastern European, cities. The study which incorporates the general context of a postmodern city analyses the shifts of meaning of the traditional bohemian ethos in the face of such processes as globalisation. postindustrialization and the emergence of the new media. I intend to analyse the process of the inclusion of the socially marginalised spaces to the mainstream, distancing at the same time from the concepts of Florida's 'big morph' or Brook's 'bobos'. While investigating the relationship between bohemian artistic tradition and consumer culture, I want to focus on the spaces of articulation which allow the subversive power of the traditional bohemia to take shape. Following Lefebvre's statement that space is a social product. I intend to characterize spatial practices that constitute artistic space and reconstruct the reasoning behind them. My research will be concentrated on the chosen cultural hubs of Polish and, I hope, neighbouring countries' cities. It is my intention to witness and describe the unique processes as they are happening now. What is more, I would like to indicate the increasing role of the arts in the production of economic capital as well as comment on the experience of a postmodernist city which is additionally influenced by its socialist legacy.

Keywords: *neobohemia, creative class, postmodern urbanism, spatial practice*

1 Introduction

Nowadays, it has become impossible to speak about cities without recognizing the escalating role of the arts and artists in the production of urban space. The growing importance of the arts in almost every aspect of city life has been noted by many scholars of urban studies and sociology of culture1. Although the tradition of an artist in the Western city can be traced back as far as the 19th century, a notable shift of that notion should be taken into consideration if we want to consider it in the postmodern city context.

The following research is aimed at the critical review of new processes that take place in the contemporary urban environment and which, as I believe, draw from an augmenting imaginary of the artist in the city space. Following Richard Lloyd's claim about "bohemia going global", I want to investigate the impact of artists on the cultural capital of Central and East European cities. My choice of the cities is inevitably pragmatic but at the same time there is a methodological background behind it. Lloyd's observations in Chicago's artistic district Wicker Park, although inspiring and insightful, cannot be uncritically transferred into European context as they are deeply rooted in a specific historic and social context of American urban experience. My intention is to take a closer look at the traditional and budding artistic neighbourhoods in the general context of postmodern city. One of the basic questions is the complex relation between art and commerce which is the underlying theme in Richard Florida's popular notion of the creative class. In my research I want to investigate how well the concept of neobohemia inscribes into what is a thoroughly political and economic project of creative cities. Does the tradition of the artists in the city with all its social order-disturbing connotations undermine or support the logic behind cultural economy? How much subversive power, if any, is left of what Balzac rebuffed as "the vagrant students of the Latin Quarter" (Snyderman and Josephs 1939: 187)? The next aspect of my research is focused on the city as a condition sine qua non for the existence of neobohemia. It is the city that constitutes the primary environment for the artistic practices which, in turn, give meaning to particular urban spaces. If we assume Lefebvre's statement that space is a social product, we will be able to characterize spatial practices that constitute that space and perhaps reconstruct the reasoning behind them.

2 Literature review

The reconstruction of bohemian tradition in connection to the development of modernism will be based on the 19th and 20th century literature, as well as critical research of Benjamin (1999), Bourdieu (2007), Nicholson (2003), Wilson (2003), Kleinerman (1932), Franck (2001), Marve and Moss (1946) among others. My research draws also from Cesar Graña's (1990), Mary Gluck's (2005), Elizabeth Wilson's (2003) and Jerrold Seigel's (1986) comprehensive studies about Bohemia.

¹ E.g. Sharon Zurkin, Doreen Massey

The literature of the period is vast and therefore I will consider a few most prominent authors such as Baudelaire, Gautier, Murger, Poe, Boy-Żeleński.

Postmodern urban context of neobohemia will require a theoretical approach and therefore the following authors will be taken into consideration: M. Dear, E. Soja, S. Flusty, M. Gottdiener. My research will also refer to the studies on urban experience and strategies of urban perception of Lefebvre and de Certeau among others. The context of creative cities and cultural economy will take into account such authors like R. Florida (2002), Ch. Landry (2000), E. Glaeser (2011), A. J. Scott (2000).

3 Neobohemia and the city: methodology

At the beginning I feel obligated to stress the fact that the concept of neobohemia is elusive and can be quite hard to define. Like Bauman's liquid modernity, it is profoundly mobile and inconstant. Lloyd (2006: 50) admits that "the boundaries of bohemia may be porous, both spatially and conceptually". On the other hand, it comes as no surprise that in most contemporary cities certain 'artistic districts', can be distinguished from the 'ordinary' urban environment. These neighbourhoods attract more and more scholarly interest as being the hubs of symbolic as well as material capital. In Lloyd's view, the experience of living and working in such places relies strongly on the bohemian ethos.

The practices of everyday life and work are perceived, by both the inhabitants and the outsiders of the neighbourhoods, in Lloyd's claim, as a continuum of certain bohemian symbolic imaginary. I believe that this statement requires further analysis in a local context of case studies I intend to undertake. Are there any specific spatial practices which can be said to continue bohemian tradition? If so, to what extent are they inscribed in the places themselves? Answering these questions, in my opinion, will entail historic reconstruction of the local bohemian tradition. As bohemia is here recognised both as a material space as well as a state of mind (Lloyd 2006: 50), this will prove helpful, I believe, in an attempt to analyse how conscious the relationship with past artistic tradition is in the minds of the contemporary inhabitants of the neighbourhoods.

The modernist city was inseparable from the bohemian ethos. It provided the creative milleu, was the reference point of the arts; it finally constituted the ultimate space for creative expression and bohemian life. The cultural phenomenon of *bohéme* is strongly connected with the aesthetic practice of modernism, a link which has been acknowledged by many historians of bohemia. It is worth to remember that the idea of 'living like a bohemian' has always been a social construct whose origins can be easily traced from the works of Théophile Gautier, Henri Murger or Honoré Daumier just to name a few. The images of bohemian archetypes are abundant: from Baudelaire's *flâneur* to Mailer's hipster and Kerouac's beatnik. Bohemian experience always emphasised the urban underworld as the 'authentic' urban experience. The aestheticization of the decay still constitutes a part of neobohemian practices. However, the context of the modern space which gave birth to the bohemian has changed dramatically.

One cannot diverge from the context of processes which shape modern urban condition. Postindustralisation, globalisation and the development of new media are responsible for the shift of the meaning of bohemian tradition. The advantages drawn from the industrial morphology of the urban space undoubtedly have an impact on industrial production. The arts and artists no longer occupy the marginal space of city economy. Nor do they oppose the capitalist order all together, as shown by Daniel Bell. On the contrary, they encourage its growth. This shift from the margins to the privileged centre of cultural production and consumption is one of the notable changes which take place in the new understanding of the bohemian concept.

The radical break from the previous cultural, economic and social trends denoted by the *post*- prefix2 which has appeared in the field of urban theory must, in turn, be reflected in the notion of *bohéme*. Neobohemia spans over two disjointed spheres – its deep reliance on the history of industrial neighbourhoods and the decentralised, fragmented reality of postmodern metropolis. The gap between the reliance on a fix and solid place, like e.g. a factory, and the elusiveness of activities characteristic to the virtual and the innovative was shown for example by Harvey Moloch. It is not my intention to analyse in this paper the complex entanglements between postmodern urban condition (Dear 2000) and the persistence of the cultural phenomenon of bohemia. I hope to develop this theme in the course of my doctoral research. Here I would like to pinpoint the areas which I think are worth a closer study.

As culture and the city are intertwined and affect each other reciprocally, the radical postmodern change must be reflected in what is one of the most notorious and alluring urban ethos. One of such fields of interest is the change in the production and work pattern triggered by global processes and regulated by dynamics of the high mobility of capital, which results in "evad[ing] long-term commitment to place-based socioeconomies" (Dear and Flusty 1998: 61). The allure of the hip neighbourhood and the vagabond ethos of bohemia attract the most innovative and digitalised sectors of economy which are highly mobile and place-independent. This contradiction between the virtual work pattern and a longing for a fixed place indicates the discrepancy of urban experience. Another perspective which neobohemia can be juxtaposed with is its role in what Dear and Flusty call a commudity (1998: 63), i.e. a commodified community designed for need of the cybergeoisie. Can neobohemia, with its hip glamour, insistence on authenticity, the aestheticization of crime and the notoriety become a valued commodity, a simple exploitation of a specific way of life? Or, being a result of memetic contagion (Dear and Flusty 1998: 62), is a space from which new oppositional identities and political agendas arise? These are just a few questions which require a thorough consideration.

The role of neobohemia in the production of cultural capital touches upon the issues raised by Richard Florida. His 'bohemian index' is another piece of evidence for the existence and the rising role of neobohemia in urban economy; however, Florida (2002) himself rejects this notion arguing that we cannot speak about bohemia in a traditional sense as artists do not feel alienated from the economic order. The distinction between Lloyd's and Florida's concepts is quite crucial. The two ideas

² Ewa Rewers in her introduction to the philosophy of postmodern city signals the break between the post- and polis, with the postmodern urban experience and the idea of *polis* drifting away to different directions.

seem to stem from similar premises such as tolerance, innovation and creative labour; however, Florida's explanations seem to be too broad to touch upon the elusive nature of the process. I believe that research concerning artists as a resource of symbolic capital should be contextualised to the specific urban environment which they occupy.

My focus on creative industries and creative cities projects does not mean that I am uncritical about it. On the contrary, it is my belief that complex power apparatus stands behind every cultural venture supported by the state and other control centres and to belittle this fact would be both unwarranted and harmful. In this respect I follow Gramsci's theory of cultural hegemony and the political inquiry of Chantal Mouffe and Ernesto Laclau. What is more, I want to focus my attention on the processes of appropriation and exploitation of certain aspects of culture of the minorities and the underprivileged as these marginalised spheres have always been the natural environment for the bohemian artists. In the course of further research I want to investigate the ways in which the search for urban 'authenticity' supports the logic of accumulating symbolic capital of the cities and what role neobohemia plays in this process. I see two ways of dealing with that issue – one, proposed by Matteo Pasqinelli and his notion of *factory of culture* where styles of life become commodities for sale and David Harvey's analysis of the monopoly rent and the 'spaces of hope' which allow the opposition and alternative movements to take shape. Both approaches, I believe, can be employed for explaining the complex role that neobohemia plays in the idea of creative cities. What is most interesting for me in this respect is the disjunction recognised by Harvey between the capital's search for homogenisation and authentic experience. The accumulation of the rent requires certain diversity and therefore it is crucial for the city states to support different, even if 'transgressive', cultural practices. Perhaps this could be the exact space where neobohemia can articulate its critical and subversive potential.

4 Conclusion

I believe that my research will contribute to the better understanding of new processes that we nowadays witness. It seems that the new city context coerces new ways of urban perception. Postmodern city has made the iconic figure of traditional *bohéme*, the flâneur, obsolete. As Ewa Rewers (2005: 6) reminds us, the postmodern urban space experience is no longer dominated by the pedestrians, the spectacle, the pleasure of voyeurism and the danger and the crime on the city margins. The experience of the city has become fluid, dynamic and self-reflexive which, partly, is the result of the appearance of new virtual spaces and digitalisation. This shift, I think, can be observed in the constant changes of the dynamics of urban space, especially in the newly established artistic neighbourhoods as they become the primary spaces of cultural production.

Lloyd's research in Wicker Park was conducted after the neighbourhood's decline. I have an outstanding opportunity to look closer at the processes as they are presently happening. With Polish, as well as Eastern European cities, undergoing radical economic, political and social changes and dealing with socialist legacy, the findings of my paper seem to be quite innovative as no research concerning the cultural ethos of the artist and its role in the production of postmodern urban space and symbolic capital has, to my knowledge, been proposed for this part of Europe.

In my research, I am most interested in the individual interpretation of the city by the users of the space and their representation of the city image. The analysis of the spatial practices in the neighbourhoods would entail an epistemological perspective which, after de Certeau, we could call a walker strategy. Being aware of the fact that such a perspective offers a fragmented and subjective view on the process under investigation, I believe that microhistories are an inspiring tool for the reconstruction of what I call neobohemian self-consciousness. The interviews with the insiders of the neighbourhoods and members of the artistic and intellectual cliques will be supported with the analysis of texts and artistic works as the products of a unique cultural ethos of *bohéme*.

References

Bauman, Z. (2006) Liquid Modernity, Polity Press, Cambridge.

Bell, D. (1978) The Cultural Contradictions of Capitalism, Basic Books, New York.

Dear M., S. Flusty (1998) Postmodern Urbanism. Annals of the Association of American Geographers. Vol. 88, No. 1, pages 50-72.

Dear, M. (2000) Postmodern Urban Condition, Blackwell, Malden.

Florida, R. (2002) The Rise of the Creative Class, Basic Books, New York.

Gluck, M. (2008) *Popular Bohemia. Modernism and Urban Culture in Nineteenth-Century Paris,* Harvard University Press, Cambridge.

Harvey, D. (2002) The Art of Rent: Globalisation, Monopoly and the Commodification of Culture . *Socialist Reader*. Vol. 38, pages 93-110.

Josephs W., G. Snyderman (1939) Bohemia: The Underworld of Art. *Social Forces*. Vol. 18, No. 2, pages 187-199.

Laclau, E., Chantal Mouffe (2001) Hegemony and Socialist Strategy, Verso, New York.

Lefebvre, H. (1991) The Production of Space, translated by D. Nicolson-Smith, Blackwell, Oxford.

Lloyd, R. (2006) Neo-bohemia. Art and Commerce in the Postindustrial City, Routledge, New York.

Lloyd, R. (2004) The Neighbourhood in Cultural Production, *City & Community*, Vol. 3, No. 4, pages 343-372.

Molotch, H. (1996) LA as a Design Product, in Allen, S., E. Soja, (eds), *The City: Los Angeles and Urban Theory at the End of the Twentieth Century*, University of California Press, Berkeley, pages 225-275.

Pasquinelli, M. (2008) Beyond the Ruins of the Creative City: Berlin's Factory of Culture and the Sabotage of Rent . *Text presented at Consistory Talk I: The Artist and Urban Development*. Downloaded from: www.matteopasquinelli.com (5.04.2012).

Rewers, E. (2005) *Post-polis. Wstęp do filozofii ponowoczesnego miasta,* Universitas, Kraków.

Albanian Language under the Pressure of Social Development

Adelina Sula¹, Shezai Rrokaj¹

¹ Department of Albanian literature, Faculty of History and Philology, University of Tirana, Albania adelinasula@gmail.com

Abstract. As we know, English is exerting influence in many European languages. Even in today's Albanian language this influence is felt, especially in the 90s of last century following. Imported English words rushing, especially in computing, in journalism, etc. Unification is a process of mutual awareness of the totality of social values, historical, cultural and linguistic. He performed in conditions where the potential value not only remain, but return to the active part of every individual. Social changes have led the life-changing vocabulary of Eagles. bringing an unprecedented enrichment of vocabulary with new words and expressions, words tend to go toward existing passive vocabulary. Thus, every language has a vocabulary peculiar to certain historical time and is difficult to understand words that are not reflected in the dictionary. Communication is not achieved simply by word of funds under certain rules, but their combination in accordance with the disclosure of reality, according to the intention of the speakers in the circumstances of communication. This paper is part of the literature that used related to vocabulary. Especially have served these works "Dictionary of modern Albanian language" 1980, "The Albanian language Dictionary" 2006, issue of the Albanian language, philological studies, etc. Also, is exploring what changes there are between the Albanian language dictionaries.

Keywords: English, Albanian, globalization

1 Introduction

We live in an increasingly global world. There is simply matter to the world market, which is a fact of the nineteenth century. Third millennium seems to require increased accountability of individuals free to determine and understand the essence of the era of globalization.

Meanwhile inflows of foreign words are becoming inevitable consequence of the revolution in information, telecommunications and digitalization. Of nearly twenty years the Albanian society is undergoing fundamental changes, which paved the way for the overthrow of their previous political system. The consequence of these changes was the free movement of people, new possibilities of their turnover, which grew and expanded communication with the world's Albanians, said their contacts with the languages and cultures of other peoples. Of course, not only among Albanians, but now everywhere intercultural and linguistic relations are displayed with a momentum unprecedented dynamism. Tremendous technological progress, especially that of information technology (internet, satellite television, mobile telephony, etc.), But also expanding education, transformation of English in an international language, etc. are key factors to these changes.

Meanwhile, thanks to progress, with reductions downs over the centuries, have reached a new target, with numerous opportunities for increasing the level and quality of life, with a value system that make possible the coexistence in conditions of freedom and the democratic governance at local, regional and global. Elaborate efforts to developing it, putting the individual at the center as a microcosm, social and ecological environment in which he lives and will have to live in the era of globalization. Subordinated to this object, as "brick" construction estimated wordcommunication-information. Language is not a closed system, no rigorous scientific discipline, but a culture that is derived from the whole process of education, education, dialogue and communication, the media and experience, without excluding or dialects.

2 Literature Review

Year 1990 has already emerged as the biggest year of the overthrow, the overthrow of the "communist empire". Inversion, even though objectively it had become inevitable, not just SICC was expected to happen happened, as the twinkling of an eye, where the "castles" fell one after another, if they had been sand or cardboard. So it was a system, mounted on an era of utopia, the illusion of real desire for rapid development, which operated a moral principle, a policy of an ideology and system in place, of course, everything created in the function it fell together with the system, as well as many of the texts fell, articles, books and brochures published propaganda at that time.

Changes that have occurred in Albanian language lexicon in recent years, as observed in the general vocabulary in both the terminology, they are quite pronounced in the field of Politico-social vocabulary. So, this lexicon is characterized by an ideological character. The language of social and political life of many other areas takes the standard features of this style, everything starts to enter the stereotypes of appropriate policy and obtain a whole layer of hardened phraseology. From this thematic division appears that all the vocabulary of this field is essence of which the Albanian language today includes terms related to political and social organizations, in general, the State and philosophy.

In this regard, the overthrow of the totalitarian system was a kind of catharsis for all but a few more to that small group of intellectuals' ideological and theoretical front, for the chance to reflect, to see their previous activity as a critic period and commit to a renewable policy and thinking otherwise.

But, until the time comes the "paradise on earth", the priority should have the spirit of sacrifice and totalitarian attitude towards life and people can have quite simply the fate of two heroes of novels and a missing process, where the guilty punished indiscriminate death.

Removing geopolitical obstacles that impede the movement of people, hence, goods, ideas, social and cultural relations, made the Albanians leave their enclaves, where he had set the historical injustice. Along with them the culture, language, institutions, ways of perceiving the same world and the formation of the national language awareness. Free Moving people within the enclave, step by step, is "warp" dialects and dialects, at least in the classical sense of these concepts, giving a new view of the Albanian language in Atlantis.

During these two decades, the Albanian vocabulary has developed and expanded in a conscious and organized, is continuously enriched with a progressive, becoming more and able to respond to demands of economic development has paved social-country. All these cannot bring and not bring significant changes to the vocabulary of our language. This has made it possible for the Eagles put the dough in a much wider movement, by increasing contacts. This intensive development and enrichment of our language gives us the full right to talk to a new level of vocabulary development, evolutionary character development.

In such circumstances it is completely understandable new Albanian situation where the needs of all kinds, but penetrate with dozens words from other languages are emerging every day, especially in print. This pressure is inevitable in the early stages of opening to the world. Thus, there is an impact of foreign words and structure to our language, mainly "English words" and "Italian words". This relates to new relationships that are created in the Albanian language today between its vocabularies and foreign. These words have in common with other languages, are making them international cannot take the place of dough words of the language by creating new words. Meanwhile, the introduction of words and constructions of foreign European languages, especially in official discourse (in the field of science and technology, in journalism, in literature translated) continues to be quite sensitive. Such words are the most scientific terms, such as chemistry, physics, mathematics, biology, philosophy, etc., used in all languages of the world.

The scope of a word from one language to another conversion to nationalism, in most cases determined by the weight of the word, the value, importance or role of social phenomenon, the object, the notion that this word is marked by . They mark a phenomenon notion that cannot express or to record by means of the Eagles. Acquisition and use of borrowings in Albanian is a way of enriching its vocabulary, but other ways of getting rich. Each of `international words ,,in Albanian language, regardless of origin, took place the firm in its vocabulary, has undergone phonetic and

grammatical system of our language, which is written and pronounced by the phonetic value of our alphabet.

Most Albanian" international words` special character, are terms or words special place among the leading international words occupy those related to Politico-social life and various science and technology, where we can mention the words: "aerobi, alkin, aluvionar, asteroid, digjital, diktofon, disketë, ekosistem, ekuinoks, elikzir, empirist, isobar, katalizë, kompaktdisk, kompjuter, kursor, maus, radiofoni, resonator, rrjetëzore, rrjetinë, saraçineskë, semiotikë, silenciator, skaner, specie, spektroskop, sprucator, shpatullore, termoizolues, tokujë, tokurë, valvul, vektorial, buton, celular, citric, epruvetë, etnolog, etnologji, faks, fishë, hidrosferë, historiografi, informatikë, kamikaz, klonoj, magnetoskop, masmedia, master, monom, oksiton, oktaedër, orbitale, orientalistikë, orografi, printer, radioamator, radiotelegrafi ,telefonia, telekamerë, telekomunikacion, toponomastikë, video, videokamerë, videokasetë, videolojë, videoregjistrues, videotekë, etj." [2]

3 Proposed Methodology

The degree of similarity, and consequently the degree of ease, to recognize and identify international words changes. To determine whether the identification of these words more essential is to have similarities to regular or consistent theme sounds or root, for it is they are primarily associated with giving meaning to the word, which may very well ' analog call and international semantics.

Enrichment of the target language is legal process, in which participate the act every day different layers of people, starting with the common people, artists, writers, military, science and the various specialists of production. The more they develop, so they need new product names, word, phrase or way of speaking that respond better, more precision notions of ideas, which express more strongly with fully life and man's spiritual world.

It is known that the borrowing of words and terms of socio-political vocabulary, as well as rental of general vocabulary, is conditioned by linguistic factors (the needs of the language), and several factors extra linguistic (end geolinguistic sources and Sociolinguistics), including the attitude (position) of different social strata to foreign words.

Of course, the rate and extent of use of loan depends on the style and timing of events or phenomena described the presence or absence of synonymous words in Albanian for them, or linguistic culture and artistry of one who writes or uses words.

However, our language, throughout history, but especially in the context of globalization (in the present political circumstances, geographical, economic and social) cannot remain without influence from foreign words, especially internationals. In Albanian language have penetrated these years and a host of words from other languages, a large part of which because of its needs for the named objects and new concepts, but also to express new meanings or stylistic connotations, to enrich the variety of expressive tools, etc..

These loans to so-called "loans required" can be considered already part of the vocabulary of our language, regardless of their origin. Remember that the eagle, for

many objective reasons, is now a language quite open to many innovations other languages and cultures, which, however, should not make us to reduce the effort to preserve its identity, but not to stop processes that are complicate this identity. Modern eagle should be a tool of modern culture, which not only preserves the Albanian identity, but the trend appeared sound European cohesion.

Meanwhile, everyone has a right to doubt and question: "Today, at the dawn of the third millennium, the era of information technology development, the" expansion "of English and its transformation into a" global language ", the contacts not seen between languages and cultures, is it possible that it covered all names of new arrivals, which flows every day in the media and, of course, not only to him?"

There is no doubt that information technology has become today an integral part of everyday life to many of us, and together with, the English, the language of its instrumental. No one makes more impression, as he reads a manual slip whatsoever, facing it with dozens of English words. This is because English has consolidated its dominant position as the instrumental language of informatics. Thus, the underlying fund of vocabulary in many languages, information from the Albanian and Italian to Japanese -consists of English words, and not only. However, this does not invalidate that no efforts are made to become more systematic and organized for translation in Albanian language and standardization of a number of terms.

Today more and more modern society is feeling the need for new language means more processed expressive, whether terms, whether general vocabulary words. Deepening and proliferation of knowledge, its application in manufacturing and services, specialization of labor, further differentiation of institutions, etc. today constitute the main impetus for creating regulations. However, even today, in the era of globalization of informatics universality, linguistic planned interventions are not always made simply and exclusively with the intention to serve communication.

Even the eagle standard, like many other languages, admitted quite borrowings, the need for labeling and the need to make communication more efficient. This development and enrichment of vocabulary is a way of reflecting the interaction of our language with other languages, and requirements of society to integrate the culture of our people with other peoples, but without forgetting care to preserve and protect the national itself, thus creating a tradition of linguistic awareness raising up the people to protect and maintain the identity of our mother tongue, as a fundamental element of the nation.

Nowadays, more and more people are talking about the role or influence of "big" the borrowing of words and their meanings of international languages in terms of globalization, as the main way of enriching the vocabulary, including political lexicon social. Socio-political vocabulary of today's Albanian language has developed and continues to process its content, based on which stands its dialectical relationship with the contemporary reality of our society, as well as concepts and worldview of today's holders of Eagles, who want to see themselves on the same level social, political, economic and cultural development with that of other peoples in Europe and beyond.

Of course vocabulary is not the only area where the impact of foreign words is today present in Albanian: there he is simply the largest and most visible than anywhere else. Other influences from English, though not always useful, mainly restricted to press, the eagle is experiencing at least twenty years. The difficulty lies not only in dealing with foreign words, but in fact, that in present conditions of development of our language are the boundaries between domestic and foreign element is very mobile and unstable, a significant and complex simultaneously.

We must say that today there is a tendency in all languages of modern advanced societies to restrict or to set aside traditional use of certain words, which designate various, objects technique. For this reason, are put to use other words, until recently unknown or little-known and is encouraging the use of existing words that are neutral names.

The danger that today threatens the Eagles are not words that come from languages other than limiting its use in front of other languages "large", as long as not used where it belongs.

Should I continue to use this word (foreign) in Albanian or can be replaced with a domestic question? It is important to say clearly that the use of this word, as well as many words from other languages, should be judged in a broader context than just language. The issue of loan should be evaluated in a cultural context, even a culture, like that of today, along with the preservation of national character, increasingly gaining, too, features and international elements. In Europe today very dense traffic of people, ideas and objects of intercultural mixing highly favored.

Except that, in all European languages have existed historically and continue to exist some areas of the so-called cosmopolitan vocabulary, in which there are many words with the "international" word spread. Such are the words associated with p.sh clothing and fashion, art, science and technology, trade, international politics, etc. Using a series of loans already embedded in the Albanian should be sought, in my opinion, once the negligence or rashness of reporters he translators; the time the desire to attract the attention of the reader, therefore, be a kind of snobbery of the scribes. However, in all these cases this way writing should be avoided.

Are they welcome in Albanian literature these new forms of words? Of course not, as long as these variations are simply preferences of individuals or narrow layer `social words` no wider coverage.

In every era of human history has a close connection between literary language and the language policy of the company. Language, as the basic feature of any ethnic community as the most faithful reflection of her life, is the most important legacy of social tradition that connects generations not only of an era, but also a continuous chain of all generations.

In addition, as a tool that serves to establish categories of thought and conscience of mankind, as the direct reality of thought, language is on the one hand a necessary condition for the creation and development of culture, but, on the other hand, the production of its irreplaceable tool for the development of career advancement for all people. This shows that it is subject to the laws of development. For this reason the language is kept conscious attitude, active and worked for the regulation of its organization in the different member levels.

To meet the speakers of Albanian words or terms that did not have before, to meet all cultural needs of the people, the Albanian language has had and will have to constantly increase its means of expression, words and new meanings, to screening all her best. It is also seen in the 2006 Dictionary of the Albanian language, which provides clear mirror's treasure in our time. This dictionary float for the first time a small quantity of unknown words or less known, collected from various sources.

4 Conclusions

The above study enabled us to analyze and interpret developments and changes its quality, reaching the following conclusions:

- During the last two decades, the vocabulary of Albanian language is extended to develop an organized, is continuously enriched with a progressive, responding to requests that I have added socio-economic development of our country. This intensive development and enrichment of our language gives us the full right to talk to a new level of vocabulary development.
- The eagle standard, as other languages, admitted quite borrowings, primarily from the need for labeling, and secondly, the need to make communication more efficient. Albanian-language takes words from other languages, necessary to communicate in the field of science and technology, art and culture.

We concluded that unity is achieved when all nationalities aware of the totality of values that we belong. One of the greatest values to its language, part of national identity.

References

A group of authors, (1980), Dictionary of Modern Albanian language

A group of authors, (2006), Albanian-language Dictionary

Rrokaj, Sh. (2009), Issue of the Albanian language, Tirana

Steel, M, (2006), Standard Albanian Affairs, Tirana

Lloshi, J. (1999), Stylistics of the Albanian language and pragmatic, Tirana

Shkurtaj, GJ (2008), How to write Albanian, Tirana

Samara, M (2008), About Politico-social vocabulary of Albanian, Tirana

Jorgaqi, K (2008), desk The Eagles

Duro, A (2009), The term and the word in Albanian, Tirana, 2009

Kole, J (1984), On international words in Albanian, SF

Kostallari, A (1972), in the way of drafting normative vocabulary of Eagles, SL.II

Washback of State Matura Examination on teaching

Jaup Zenuni

Foreign Languages Center, "Spiro Moisiu" Defense Academy, ja12ze@yahoo.com,

Rruga e Dibrës, Tirana, Albania.

Abstract. Washback is a phenomenon that influences many aspects of language education in schools. Its complexity has attracted the attention of many scholars. It is one of the domains of the Applied Linguistics that gained prominence especially in the late 80-ies when a great number of researchers focused on different approaches to this phenomenon. The influence of public exams on teaching, learning and curriculum is otherwise described as washback. Washback on teaching is thought to be the area where many scholars have focused their attention. Hughes (1988), Alderson & Wall (1993), Shohamy et al. (1996), Cheng (1997), Cheng, Watanable & Curtis (2004), Green (2007), Ying (2010) and many other prominent experts of the field have greatly contributed in studying the washback effect on teaching and teachers.

Although teaching is one of the most examined aspects, there is still a great need for fresh contribution. This paper is an effort to examine some of the effects of washback of State Matura Exam (SME) on English teaching in the context of the Albanian high schools. Through some empirical data collected from 109 teachers and 401 students from different high schools by questionnaires prepared and administered for this purpose, the paper examines teaching aspects mostly affected by washback of SME.

Keywords: Washback, State Matura Exam, CEFR, teaching, test.

1 Introduction

State Matura Exam is a system regarding final exams that young adults take in the end of their high school education in Albania. It was created by the Albanian Ministry of Education and Science (MES) in 2006. The main purpose was to improve the quality of education in high schools, avoid corruption, lack of control and favoritism during the process of students' university entrance. According to policy makers this new examination system for high school leaving had four main goals: 1. Certification -

issuing a maturity diploma, which certifies that the student has finished the high school education; 2. Selection-selecting top students to enter university based on the scores; 3. Monitoring-monitor the implementation of education standards and ensure the quality of education; 4. Informing-inform the policy makers and stakeholders for the quality of high school education and for students performance. SME consists of two compulsory exams (written Mathematics and written Literature) and two other exams. These last two exams are chosen by the student from a pool of subjects (Chemistry, Biology, Physics, English, History-Geography and Sociology-Economy-Philosophy). What two subjects they choose depends on their university preference. Each university's faculty chooses its own preferred profile (scientific or social) and its own preferred subjects.

Since the very beginning of this new system English has been one of the preferred subjects by students who take SME. For instance, for 35% of the students that took SME in 2011, English was their preferred subject. According to the policy of MES, from 2013 English is foreseen to be a mandatory examination.

Therefore, it is natural to suppose that the introduction of SME in English language has had some repercussions since this subject is, for the first time in the history of education in the country, considered as one of the most important. Hence, some questions arise: Has SME in English had any impact on teaching? If so, what is the level of this impact? Which of the teaching aspects are influenced more? How is the examination reflected in the classroom? This paper will try to address these issues.

Most of the scholars engaged in research on the influence of testing on teaching, learning and curriculum have called it washback. As the focus of this paper is going to be the influence of testing on teaching, it's necessary to explore the body of research with reference to it.

2 Literature Review

Washback¹ or backwash has been defined as 'a part of the impact a test may have on learners and teachers, on educational systems in general, and on society at large' (Hughes, 2003). There is no doubt that high stakes and decision makers have used language examinations to lead to changes and reforms of the system. Speaking about the use of examinations Bachman (1990) states that the only and most important reason of conducting tests in foreign language and interpreting their results is the

¹ The term 'washback' is more universally used in the field of applied linguistics to refer to the influence of tests in language education (Alderson & Wall, 1993; Cheng & Curtis, 2004, Hawkey, 2006; Green, 2007; Shih, 2009). This is the reason this term will be used in this paper.

purpose they are supposed to serve. Almost all the research work has demonstrated that public exams produce the greatest amount of washback. Messick (1996) highlights the idea that to be considered washback, good or bad teaching has to be 'evidentially linked to the introduction and use of the test.' Shih (2009) thinks that researchers have paid most of their attention to the washback of tests on four domains of teaching practice:

- 1. Content of teaching,
- 2. Teaching methods,
- 3. Assessment methods,
- 4. (More broadly) overall teaching style, classroom atmosphere and teachers' feeling towards the test.

Wall and Alderson (1993) developed a basic framework for the Washback Hypotheses, which simply assumed that tests have impact on teaching. Six of their hypotheses, out of fifteen, directly address teaching. The results showed that washback on the content of teaching existed; however, washback on teaching methodology was not observed. This finding is also seen in Chen's (2002) study of the impact of a public exam on English teaching in Taiwan, which revealed that the test 'may dramatically change the content teachers teach, but not the way they would teach.' In general, many authors agree that language tests seem to be effective in bringing about changes on teaching content but have only limited apparent success on teaching methodology (Alderson and Wall, ibid. 1993; Shohamy et al., 1996; Cheng, 2005; Luxia, 2007).

In contrast to the universal presence of tests washback on the content of teaching, teaching methods were not changed by tests to a great extent. They were shown to be altered in some studies (for example, Alderson & Hamp-Lyons, 1996; Ferman, 2004; Hawkey, 2006; Stecher et al., 2004), but not in others (for example, Cheng, 1997, 1998, 1999; Alderson & Wall, 1993). Lam (1994) after surveying a number of teachers that had worked both with the old and new syllabuses and another group of teachers who worked only with the new syllabuses, came to the conclusion that the first group was much more test-oriented than the second one. Hence, he thinks that: "The challenge is to change the teaching culture, to open teachers ' eyes to the possibilities of exploiting the exam to achieve positive and worthwhile educational goals." Chapman & Snyder Jr. (2000) also question the extent to which high-stakes testing influences teachers' classroom methodology. Citing a general education study from Uganda (1997), which found that changes made to a national examination did not have the desired effect of encouraging teachers to alter their instructional practices, they suggest that 'it is not the examination itself that influences teachers' behavior, but teachers' beliefs about those changes'.

In regard to washback on assessment methods, suffice to mention that assessment is located in both a narrower educational and a broader social context, and the

involvement of not only direct participants (teachers and students), but also indirect participants including those from the broader social context (education personnel, parents, media and the community) adds both greater importance to the washback phenomenon and different degrees of complexity. Alderson and Wall (ibid, 1993) have also underlined the close connection among testing, teaching, assessment and learning. The most of washback from the tests is produced in the triangle teaching-assessment-learning. This implies that what is assessed strongly influences what is learned. Studies have shown that no greater impulse for learning exists than assessment (Frederiksen, 1984), with some authors even stating that innovations in education are destined to fail if they are only implemented in the instructional process and not with a concomitant innovation of assessment (Cizek, 1997).

It can be concluded that different language tests, especially public exams, induce a wide spectrum of washback on the content of teaching, teaching methods, and assessment. Many variables were reported by empirical research to affect the degree of washback on teaching which include contextual variables, test variables and teacher variables.

Teacher variables consist of teachers' abilities in the language they teach (Luxia, 2007), teaching experience (Watanabe, 1996), teacher training processes or levels of professional training (Green, 2006; Shohamy, 1993) teachers' beliefs about effective teaching, learning and test preparation (Alderson & Hamp-Lyons, 1996; Green ibid, 2006; Wall & Alderson ibid, 1993; Watanabe ibid, 1996), learning experience (Watanabe ibid, 1996), teachers' concerns for students' proficiency levels (Watanabe, 2004), the degree of teachers' familiarity with a range of teaching methods (Watanabe, 2004), teachers' perceptions of test importance (Shohamy *et al.*, 1996), perceptions of the test qualities, the degree of teachers' commitment to the teaching profession (Wall & Alderson, 1993), and teachers' willingness and capability to innovate (Wall & Alderson ibid, 1993).

All these variables demonstrate that washback is a complicated phenomenon. To better explain the variables of the washback of tests on teaching, Burrows (2004) analyzed qualitative data from classroom observations to investigate teachers' reaction to a newly introduced test. She found that their reactions to the test followed certain patterns. Instead of a single, uniform response there were different responses. Therefore, she pointed out that a new test, interpreted and shaped by teachers' beliefs, assumptions and knowledge, would lead to patterns of reaction in teachers' instruction. The curriculum innovation model, according to Burrows, established its basis on the concept that washback was a form of educational change, so behavioral models propounded for other educational changes could also be adopted to explain washback.

3 Data and Methodology

3.1 Participants

A group of 109 teachers (99 females and 10 males) from 56 high schools of the country and 401 students from grades 10-12 of 17 high schools participated in this phase of research. This phase focused on data collection through questionnaires prepared for this purpose (three for teachers and one for students).

The age of the surveyed teachers varies as follows:

Table 1: The age of the teachers.

20-30	31-35	36-40	41-45	46-50	Over
years	years	years	years	years	50
55	35	4	9	2	

The working experience of the teachers is:

Table 2: Teachers' working experience.

Less than 5	6-10 years	11-20 years	More than 20
years			years
43	36	19	11

Besides 9 teachers who did not prefer to answer the question on education and degrees the others hold degrees as follows:

Table 3: Teachers' education and degrees.

Bachelor	Professional Master	Master of Science
Diploma	Diploma	Diploma
38	43	19

Students who were the target population of our research are all from the high schools of four districts in Albania (Shkodër, Fier, Tirana, Elbasan). Among them 109 study in the 10th grade, 98 in the 11th grade and 194 in the 12th grade.

3.2 Instruments

The instruments used for this phase of the study were four questionnaires designed according to close-ended types of questions (multiple-choice, categorical, Likert-scale, ordinal and numerical). The first questionnaire addressed teachers' perception on SME. The second questionnaire covered issues related to professional growth of English teachers. The third questionnaire aimed at gathering information on teachers' knowledge on Common European Framework of Reference on Languages (CEFR).

The fourth questionnaire addressed students' perception on SME and its influence on their learning and class work.

The design of the questionnaires was a process which required a careful work. I was aware that my long teaching career was not sufficient so I studied the related literature on the methodology of research, especially that related to the techniques and principles of proper question development. The first draft of questionnaires was carefully examined by a group of experienced teachers in my workplace. Based on their remarks several modifications were made to the original version: twenty-five items were dropped out on the grounds of being of little relevance, the wording of fifteen items was modified, and around twenty-eight items were reordered to enhance the validity of responses. In their final version the four questionnaires consisted of:

- 1. Questionnaire on teachers' perception on SME 68 items.
- 2. Questionnaire on teachers' knowledge on CEFR 9 items.
- 3. Questionnaire on professional growth of English teachers 36 items.
- 4. Questionnaire on students' perception on SME 100 items.

3.3 Procedures

Based on the Albanian law I required the support of MES to collect data from high school teachers and students throughout the country. The Secretary General and Directorate of State Matura not only expressed their readiness to support but they organized the delivery of questionnaires in different regional education offices and from there to high schools. In cooperation we fixed a deadline for the collection of the completed questionnaires. In November 2011 all the questionnaires were sent to Directorate of State Matura and they put them in my hands by the end of November.

During December 2011, January and February 2012 I processed all the questionnaires and prepared the final summative table for each questionnaire. As it is usually the case with washback studies, descriptive statistics, comparison and correlations were used to analyze the data.

However, it is necessary to clarify that not all the data collected and processed are going to be used for this paper. Only those items that are related to washback on teaching are used. The rest is going to be used in the final dissertation.

4 Empirical Analyses

Table 4 shows the descriptive statistics about the extra work and pressure exerted on teachers due to the introduction of SME. As the table illustrates most of the 11 teachers surveyed accept that in most of the teaching aspects they have been under pressure to change or forced to do extra work. Aspects mainly influenced are:

"Organizing more teaching activities that focus on SME" (5 teachers agree and 5 others fully agree); "Preparing a lot more teaching materials for English classes" (6 teachers agree and 4 teachers fully agree); "Meeting new teaching challenges" (7 teachers agree and 2 fully agree); "Making more preparation for the lesson" (4 teachers neutral and 7 teachers agree).

Question: What kind of extra work or pressure (if any) do you think SME has created on you and your teaching? Please answer the following questions by grading on a 5 point Likert scale format where: 1=strongly disagree; 2=disagree; 3=neutral; 4=agree; 5=strongly agree.

Affected aspects	Strongly disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly agree 5
Following a new syllabus.	1	3	5	2	
Doing more lesson preparation.			4	7	
Preparing a lot more materials for English classes.		1		6	4
Revising the existing materials that you have.	1		3	4	3
Employing new teaching methods.	1		4	4	2
Setting up new teaching objectives.		1	3	3	4
Meeting new teaching challenges.	1	1		7	2
Organizing more teaching activities that focus on SME.			1	5	5

Table 4: Teaching aspects affected by SME.

Table 5 illustrates the descriptive statistics related to the most significant changes made in teaching in the context of SME. As illustrated by the answers the biggest change is made in the encouragement of students for a higher participation during the lesson (among 11 teachers 9 of them strongly agree and 1 agrees). Another important change is that teachers are doing a better job to integrate language skills (5 agree and 5 strongly agree). As regards to 'teaching to the test format' and 'adapting new teaching methods' 38% of the teachers agree and 27% strongly agree.

Question: What are the most significant changes you have made in your teaching in the context of SME. Please answer the following questions by grading on a 5 point Likert scale format where: 1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree.

Changes observed	Strongly disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly agree 5
Teaching according to the test format.			4	4	3
Adapting new teaching methods.		1	3	4	3
Encouraging students' participation in class.	1			1	9
Emphasizing the integration of skills.			1	5	5

 Table 5: Changes made in teaching in the context of SME.

Table 6 shows that the 11 surveyed teachers teach the language skills and content elements of SME during their teaching practices in class. It indicates that teachers mainly focus on those language skills and knowledge that they think might be part of SME. Certainly, these data are not sufficient to create a clear picture on the influence of SME on teachers' teaching to the test, but the answers indicate that 'all teachers seem willing to go along with the demands of the exam'. It is not the age and

experience of teachers that shows how much they 'turn to the test', as 77% of the 11 surveyed teachers belong to the age group 20-35 and 88% of them have worked as English teachers less than 10 years.

- Question: Do you teach your students the language skills and content elements of SME?
- Table 6: Number of teachers that teach language skills and content elements of SME.

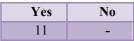
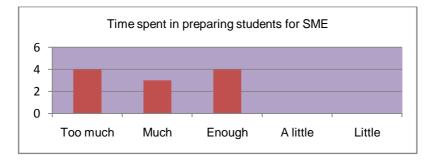


Figure 1 supports the argument expressed above (table 6) in the sense of time spent for the preparation of students for SME. As indicated, all of them spend time during their teaching to prepare students for SME. It should be stressed that only 5 out of the 11 teachers teach in the 12^{th} grade. This confirms that even in grades 10 and 11 teachers take time to prepare their students for SME.

Figure 1: Time spent in preparing students for SME.



Both arguments used above are also supported by students' perception. Among 401 students asked "Do you think that your teacher pays much more attention during the lesson to language skills, tasks and exercises that prepare you for SME", for 338 of them the response was 'Yes' and only 58 replied 'No' (only 5 refused to answer). It is significant to highlight the fact that students who replied to this question are not in the same grade: 107 of them are in the 10th grade; 96 in the 11th grade and 193 are in the 12th grade. This again brings us to the conclusion that washback of SME begins since the beginning of high school.

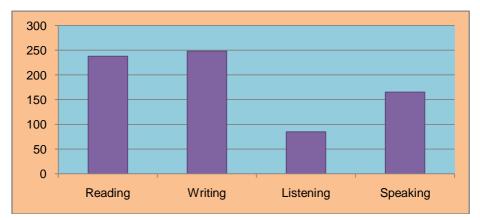
Question: Do you think that your teacher pays much more attention during the lesson to language skills, tasks and exercises that prepare you for SME?

Figure 2: Students' perception on the language skills, tasks and exercises taught by teachers to prepare them for SME.



The following figure (figure 3) illustrates students' perception on the importance given by teachers to the language skills as result of the introduction of SME. This chart clearly indicates that reading and writing, which are the skills tested in SME, have the 'lion's share'. In the total of 736 answers² (because students were free to tick 1, 2, 3 or 4 skills), 238 answers are for reading, 248 for writing, 85 for listening and 165 for speaking. As it can be noted listening is the most discriminated skill. We can easily observe that discrimination in teaching language skills brings about curriculum narrowing.

Figure 3: Students' perception on the 'share' of language skills in teaching due to SME.



In regard to the weight given to the tasks and exercises that prepare students for SME their perception is expressed in table 7. Their answers confirm that types of tasks and exercises mostly used are those found in SME. To illustrate this, the answers given by the students were compared with the SME 2011 versions A and B. In these test

² Ninety-two students refused to answer this question.

versions, 2 tasks are designed based on multiple choice techniques, 1 task is based on answers to questions on a text, 2 are based on gap-filling technique, 1 is formal letter writing and 1 is essay writing. The question addressed to the students for this purpose was: "If the answer to the question 'do you think that your teacher pays much more attention during the lesson to language skills, tasks and exercises that prepare you for SME' is YES, explain what kind of exercises she/he does more." The picture gained by a total number of 1260 responses³ (students might tick more than one item) reaffirms that grammar exercises, followed by exercises on letter and essay writing, gap filling , true-false, answers to question on a text and multiple choice are the exercises that are widely used by teachers in the lesson.

Question: If the answer is 'Yes", explain what kind of exercises she/he does more?

K	inds of exercises	Number of responses
a.	Multiple choice exercises.	130
b.	Answer the questions on a text.	144
c.	Gap filling.	179
d.	Grammar exercises.	252
e.	Matching exercises.	84
f.	True-false exercises.	149
g.	Correct the information.	94
h.	Letter/essay writing.	228

 Table 7: Kinds of exercises used by teachers.

SME has also influenced the way class tests and quizzes are designed. Teachers' responses to this question reveal that most of the techniques used in designing SME are adapted to construct class tests and quizzes. This can be noticed in the answers given in teachers and students' questionnaires. Out of 10 teachers that responded the question⁴ "How much do you use the following techniques in designing class tests and quizzes", all of them use multiple choice, gap-filling, questions on the text, matching and essay writing; 9 use letter writing and 8 use information transfer. Comparing test variants of SME for the years 2009, 2010 and 2011 it is easy to observe that the prevailing techniques are multiple choice; questions on the text; gap-filling; essay writing and information transfer (found in all test variants of the three years). The only exceptions are the use of true-false technique in the test of 2009 and letter writing in the test of 2011.

 $^{^{3}}$ Ninety-four students refused to answer this question.

 $^{^4}$ One of the teachers refused to answer this question.

Question: How much do you use the following techniques in designing class tests and quizzes?

The used techniques	Number of teachers using them
Multiple choices.	10
Gap-filling.	10
Questions on a text.	10
Matching.	10
Close exercise.	2
Open ended sentences.	4
Information transfer.	8
True-false.	6
Correct the wrong information.	7
Letter writing.	9
Essay writing.	10

Table 8: Techniques used in designing class tests and quizzes.

The above conclusion is supported by students' responses. Out of 392 students who answered the question⁵ "Do you think that SME has influenced the way your teachers design class tests and quizzes', 200 think that it has had impact on the design of these tests and quizzes while 192 think that it has not had any impact at all.

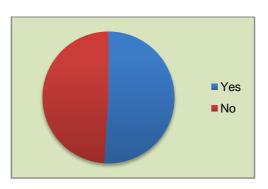


Figure 4: Students' perception on the influence of SME on class tests and quizzes.

⁵ Nine students did not answer this question.

In the following box there are some of the opinions students have expressed in relation to the influence of SME in the design of class tests and $quizzes^6$.

- Class tests contain tasks and exercises similar to those that might be part of SME.
- The format of class tests and quizzes is adapted to that of SME.
- Many tasks and exercises that are thought to be part of SME are included in class tests we have had so far.
- Class tests and quizzes have changed their format getting a simplified form of SME.
- There are cases when the scoring of class tests and quizzes is based on the same criteria used for SME.
- Teachers pay much more attention to tasks and exercises that might be part of SME.
- Now that we are in the 12th grade our teacher designs class tests that resemble SME. They contain a text with questions on it, multiple choice exercises, different grammar exercises and essay writing.

In collecting empirical data one of the foci was teachers' professional growth which is directly connected with washback of public exams on teaching. Among other factors that have also been found to affect the washback phenomenon, teachers' educational background is considered one of the most influential ones. Teachers' characteristics are of crucial importance in determining the extent and nature of test consequences. There is no doubt that among such teacher features their knowledge base in assessment or what has come to be known as assessment literacy gains a special significance. The adaptation of the techniques of SME in designing class tests and quizzes to a certain extent is related with the level of qualification of teachers. The implementation of a new curriculum as result of the introduction of SME had to be accompanied with qualitative qualification work in order to raise the assessment literacy of the teachers. In this way they would avoid copying the techniques of SME in designing their class tests and quizzes which brings about curriculum narrowing.

Data collected for this purpose show that among 48 teachers surveyed⁷ 17 of them replied that in the last 18 months have not taken part in any professional growth activity related to testing and assessment and 12 of them have participated only in one activity of this kind.

⁶ For the sake of illustration of the idea only a limited number of students' opinions are presented in this paper.

⁷ Two teachers refused to answer this question.

Question: In how many teacher professional growth activities dealing with testing and assessment have you participated in the last 18 months?

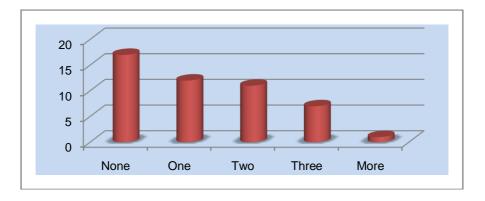


Figure 5: Participation of teachers in professional growth activities in the last 18 months.

One of the elements related to teachers' professional growth is their knowledge on the policy of MES for foreign languages. Apart from many other important major guiding goals this policy determines is that according to CEFR a student who finishes the high school should attain level B2. It is supposed that every English teacher throughout her/his work for the implementation of the English language curricula should know very well that this is a crucial goal she/he should work to achieve. Though we collected a great amount of data related to the knowledge English teachers have on CEFR, to support the arguments in this paper only the following table (table 9) is used. As shown by the data, out of 44 teachers⁸ who responded to the question below only 32 of them gave the exact answer.

Question: According to the policy of Ministry of Education and Science what language level should attain a student finishing high school?

Table 9: Teachers' knowledge on the language level a student should attain at the end of the high school.

A1	A2	B1	B2	C1	C2
		7	32	4	1

Another issue that is closely linked with the teachers' literacy on assessment and testing is the knowledge they have on CEFR which 'provides a common basis for the

⁸ Four teachers did not prefer to answer.

elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc. across Europe.'⁹ This is essential because all the curricula of the English language education in high schools of Albania bases on this document. The figure below (figure 6) indicates that there are still teachers who do not have any kind of knowledge about it. Out of 48 teachers asked whether they had any kind of knowledge on CEFR only 33 gave a positive answer while 15 responded that they did not have any idea about the existence of this document.

Question: Do you know anything about CEFR?

Figure 6: Teachers' knowledge about CEFR.



5 Conclusions

Introduction of SME has brought about radical changes in English language education of the Albanian high schools. A new language policy and a new curriculum, based on CEFR, are being implemented. Thus new textbooks published in English speaking countries like "Wishes", "Blockbuster", "Inspiration", "Prospect", "Upstream", and "Express Publishing" have replaced the former textbooks designed by Albanian authors. From this point of view SME has proved to cause a positive washback. Another positive aspect of SME washback is that it enhanced teachers and students' efforts for a higher commitment to English language teaching and learning. The use of new textbooks integrates language skills much better than before the introduction of SME. At the same time skills like writing which in the past was almost a 'forgotten skill', with the introduction of SME, became important as it is tested by this examination. The application of language standards which derive from CEFR has increased the teachers' responsibility and their teaching quality.

⁹ Common European Framework of Reference for Languages, (2001), Cambridge University Press, p.1.

However, this study revealed that there are teaching aspects which are affected by SME negative washback. As many scholars of this field have stated it is an 'unintended or unpredicted' kind of washback. This paper shows that all teachers try to 'do things they wouldn't do' unless they were under the pressure of SME. A specific finding is that not only experienced teachers are prone to teach to the test but also the young ones. The empirical data indicate that 'teaching to the test' is not typical only in grade 12, it begins since grade 10 but with the approach of SME it is more intensive. The teaching areas mostly affected by SME washback are: 1) organizing more teaching activities that focus on SME; 2) preparing a lot more teaching materials for English classes; 3) meeting new teaching challenges; 4) making more preparation for the lesson. The most significant changes made in teaching in the context of SME are: 1) encouraging students' participation in class; 2) a better integration of language skills; 3) teaching according to the test format; 4) adopting new teaching methods. Another aspect that demonstrates SME washback on teaching is time allocated for the preparation of students for this examination. The empirical data show that 36.3% of the teachers spend too much time for the preparation of students for SME, 27.1% spend much time and 36.3% spend enough time.

Some indicators that are related to the content of teaching confirm that teachers, in general, pay much more attention to language skills, tasks and exercises that prepare students for SME. 83% of the students think that teachers pay much attention to language skills, tasks and exercises that are part of SME. According to students' perception listening and speaking that are not tested in SME, are not given so much priority as compared to writing and reading. In regard to the weight given to the tasks and exercises that prepare students for SME their perception is that exercises on letter and essay writing, gap filling, true-false exercises, answers to questions on a text and multiple choice tasks are widely used by teachers in the lesson. Students' perception is also confirmed by the teachers' responses. As a consequence, this kind of negative washback causes curriculum narrowing.

SME has also influenced the way class tests and quizzes are designed. Both teachers and students admit that testing techniques used in SME are used in designing class tests and quizzes. Hence, these tests and quizzes are just a SME in miniature. This tendency prevails with the approaching of SME. 51% of the students think that SME has had impact on the design of class tests and quizzes. The adaptation of SME techniques in designing class tests and quizzes to a certain extent is related with the level of qualification of teachers. The implementation of a new curriculum as result of the introduction of SME had to be accompanied with qualitative qualification work in order to promote teachers' assessment literacy. Therefore, they would avoid copying the techniques of SME in designing their class tests and quizzes.

Although six years have passed since the introduction of SME, only some statistical analysis about the scores of the students and item relevance are conducted. It is necessary to seriously address washback issues. Negative washback can be reduced through a very well combined work among all participating actors. Negative washback in teaching may be minimized by considering all English teachers as one of the pillars of success. An updated teacher who has a solid knowledge on CEFR, the policy of MES, the new curriculum and modern teaching methods can avoid teaching to the test and aping SME. Hence there is a great need for a well organized and efficient work for the professional growth of the teachers.

References

Alderson, J.C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*. Vol. 14, 115-129.

Alderson, C,J. & Wall, D. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*. Vol.10, 41-69.

Alderson, J., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*. Vol.13, 280-297.

Bachman, L., (1990). *Fundamental considerations in language testing*. Oxford University Press, Oxford.

Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian adult migrant English program. In Chih-Min Shih (eds), *How tests change teaching: A model for reference,* English Teaching: Practice and critique, Vol. 8, No. 2, May 2009, 188-206.

Chapman, D. W & Snyder Jr. C, W. (2000). Can high stakes national testing improve instruction: reexamining conventional wisdom. *International Journal of Educational Development* Vol.20, 457-474.

Chen, L. (2002). Washback of a public exam on English teaching. *ERIC Education Resources Information Center*, 9-17.

Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, Vol.11, 38-54.

Cheng, L. (1998). Impact of a public English examination change on students' perceptions and attitudes towards their English learning. *Studies in Educational Evaluation*, Vol.24, 279-301.

Cheng, L. (1999). Changing assessment: Washback on teacher perceptions and actions. *Teaching and Teacher Education*, Vol.15, 253-271.

Cheng, L. (2005). *Changing Language Teaching through Language Testing: A Washback Study*. Cambridge University Press, Cambridge.

Cizek, G.J. (1997). Learning, achievement, and assessment: Constructs at a Crossroads, in G.D. Phye (eds.), *Handbook of Classroom assessment: Learning, achievement, and Adjustment*, Academic Press, San Diego.

Ferman, I. (2004). The washback of an EFL National Oral Matriculation Test to teaching and learning. In L.Cheng & Y. Watanabe (with A. Curtis) (Eds.), *Washback in language testing: Research contexts and methods*, Lawrence Erlbaum. Mahwah, NJ, 191-210.

Frederiksen, N. (1984). The real test bias. Influences of testing on teaching and learning. *American Psychologist*, Vol.39, 193-202.

Green, A. (2006). Watching for washback: Observing the influence of the International English Language Testing System academic writing test in the classroom. *Language Assessment Quarterly*, Vol. 3, 333-368.

Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. Cambridge University Press, Cambridge.

Hughes, A. (2003). *Testing for Language Teachers*. Cambridge University Press, Cambridge.

Lam, H. P. (1994). Methodology washback - an insider's view, in Bailey, K. Washback in Language Testing, TOEFL Monograph Series, Princeton, New Jersey, June 1999, 20.

Luxia, Q. (2007). Is Testing an Efficient Agent for Pedagogical Change? Examining the Intended Washback of the Writing Task in a High-stakes English Test in China. *Assessment in Education*, Vol.14, No.1, 51-74.

Messick, S. (1996). Validity and washback in language testing. *Language testing*, Vol.13, No.3, 246-256.

Shih, Chih-Min, (2009). How tests change teaching: A model for reference. *English Teaching: Practice and Critique*, Vol. 8, No. 2, 189.

Shohamy, E. (1993). *The power of tests: The impact of language tests on teaching and learning*. The National Foreign Language Center Washington, DC.

Shohamy, E, Smadar Donitsa-Schmidt, S.M, & Ferman, I. (1996). Test Impact Revisited: Washback Effect Over Time. *Language Testing. Vol.*13, No. 3, 298-317.

Stecher, B., Chun, T., & Barron, S. (2004). The effects of assessment-driven reform on the teaching of writing in Washington State. In L. Cheng & Y. Watanabe (with A. Curtis) (Eds.), *Washback in language testing: Research contexts and methods*, Lawrence Erlbaum, Mahwah, NJ, 53-71.

Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, Vol.13, 318-333.

Watanabe, Y. (2004). Teacher factors mediating washback. In. Cheng, L &. Watanabe, Y. (with A. Curtis) (Eds.), *Washback in language testing: Research contexts and methods*, Lawrence Erlbaum, Mahwah, NJ, 129-146.

Efficient Breach of the Contract: an Interdisciplinary Approach (Legal and Economic)

Delia-Ionela Ropan¹

¹ PhD student* - Law Faculty of Babeş-Bolyai University, Cluj-Napoca, Romania, delia.ropan@ubbcluj.ro

Abstract. This paper deals with an interdisciplinary (economic and legal) study with respect to "the efficient breach of contract theory", as developed by the economic analysis of the contract doctrine. The key question is: "Should a contract be further performed or should it be terminated in case the performance thereof is not anymore economically efficient for one of the parties?" The methodological approach used for this purpose includes analyzing the current Romanian. French and common law legal systems in order to identify whether this theory could have legal grounds. By opposition to the French legal doctrine, the Romanian doctrine consecrates only a few studies to the economic analysis of the contract. The main contributions and challenges of this study are represented by the incompatibilities and discrepancies existing between the economic approach and the legal approach with respect to this theory. The economic analysis of the contract represents a doctrine which attempts to understand contracts, being created and cultivated by economists. This study constitutes a critical overview of a jurist on the views expressed by the economists, by underlining the inconsistencies of the economic analysis as concerns the construction and application of the efficient breach of contract theory. The key finding is that my critical overview of the efficient breach of contract theory is justified by legal grounds.

Keywords: contract law, economic analysis, breach.

^{*}Investing in people! Ph.D. scholarship, Project co-financed by the SECTORAL OPERATIONAL PROGRAM FOR HUMAN RESOURCES DEVELOPMENT 2007 – 2013. Priority Axis 1. "Education and training in support for growth and development of a knowledge based society". Key area of intervention 1.5: Doctoral and post-doctoral programs in support of research. Contract no.: POSDRU/88/1.5/S/60185 – "INNOVATIVE DOCTORAL STUDIES IN A KNOWLEDGE BASED SOCIETY", Babeş-Bolyai University, Cluj-Napoca, Romania

1 Introduction

"A contract should be regarded as an instrument of co-operation, not as a confrontation of adversaries each seeking his advantage at the expense of the other. (...) It should safeguard solidarity, fairness, reliance and trust, values that everybody, also the weak party, should adhere to." (Lando, 2008, p. 187)

Analysis of contracts has always represented a challenge for theoreticians and practitioners altogether, both from an economic and legal perspective. The sensitive issues related to the life of a contract are numerous, from the beginning of the process with the enthusiasm of the first negotiations until the termination thereof, either by litigation in front of the courts of law/arbitral tribunals, by amicable settlement or by plain performance of the object of the contract. Thus, law is not only the field of interest of jurists but also of economic analysis of law). This economic analysis of the contract (as part of the construction of specific approaches/standpoints/theories. Some of them have not been previously considered *per se* by the legal doctrine or by the law.

This interdisciplinary study (legal and economic) is focused on revealing the incompatibilities existing between different views on the same subject (*i.e.*, specific institutions related to contract law). These differences concern the way in which the economists developing the economic analysis of the contract treat different aspects related to the performance of the contract and the standpoints of the legal doctrine on the same aspects. One of the theories developed by the economic analysis of contract on which we will focus herein is *the efficient breach of contract*.

The objective of this study is to answer the following questions: When economic efficiency is not met anymore (or in doubt) by one of the parties, should the contract be terminated, amended or performed further? Is the contractual equilibrium part of the economic analysis of the contract? Is the economic analysis of the contract concerned by the contractual balance/equilibrium?"

All these questions are triggered by the fact that during performance of a contract, the contractual balance contemplated by the parties upon conclusion might be endangered by certain factors. In such cases, the key question arising is whether the contract should be further performed by the parties in the initial terms agreed or if the contract should be amended or terminated.

The first reaction (particularly from an economic standpoint) would be to opt for the destruction of the contract. However, a final response to the above question is not accurate without performing an in depth analysis of the impact of the legal approach on the matter.

The main types of contracts concerned for the purpose of this study are the synallagmatic contracts (*i.e.*, the contracts imposing reciprocal obligations upon the

parties – e.g., sale, distribution, franchise, transport etc.). More specifically, our preoccupation covers the field of business-to-business contracts.

In order to address all these aspects we will start with evidencing the basic literature of great influence with respect to the topic of this paper (Section 2). Afterwards, we will list the methodology used for the purpose of this study (Section 3) and we will focus on succinct considerations regarding the contribution of the economic analysis of contract, with an extensive development of the central point address by this paper, *i.e., the efficient breach of contract* theory (Section 4). In the end we will present some conclusive remarks in connection with the topic developed in this paper (Section 5).

2 Literature review

The economic analysis of law is a movement having its origins in the United States, being firstly developed by economists at the end of 1950's (Fabre-Magnan, 2010, p. 89). Afterwards, the approach initially promoted by the School of Chicago has emerged in Europe.

The promoter of this movement is Richard Posner who has mirrored in his book on "The Economic Analysis of Law" (reedited several times since its first publishing in 1972) an analysis of the efficiency of the rules of law in all law fields, including contract law. For the purpose of this research we analyzed the sixth edition of this book which dedicates Chapter 4 to the study of "Contract Rights and Remedies" (Posner, 2003, pp. 93-143). In order to facilitate the emergence of the economic analysis of law within Europe, the preoccupation for assuring an extensive book in French was also concretized (Mackaay and Rousseau, 2008).

An economic approach of the law made from the perspective of the economists cannot be ignored by the legal environment. Thus, the jurists from France and, to a certain extent, from Romania have also been interested in acknowledging the approach of the economic analysis of the contract in order to analyze it in conjunction with the legal institutions.

The Romanian post-revolutionary legal doctrine has not consecrated extensive studies to the topic of the economic analysis of the contract (and of the economic analysis of law, in general). There are however brief notes in Vasilescu (2003, pp. 49-51; 2007, pp. 118-122; 2012, pp. 248-251) and Pop (2009, pp. 79-91).

By opposition to the lack of interest of the Romanian legal doctrine on this topic, the French legal doctrine (Cathiard, 2006; Fabre-Magnan, 2010) has broadly analyzed from a legal perspective the contribution of the economic analysis of contract, by revealing especially the main incompatibilities and inconsistencies of the approach taken by the economists.

A perfect illustration of the different visions of the economists and of the jurists on the same matter is represented by the so-called *the efficient breach of contract* mechanism on which we will focus herein below. The large development of the economic analysis of the contract (and of law, in general) at the international level justifies, from our perspective, the need for a more in-depth analysis of the topic in the Romanian legal doctrine as well. This study is intended to this aim. In addition, in opposition to the views already expressed by the Romanian legal doctrine which are rather descriptive than critical, this study focuses on the discrepancies existing between the thesis developed by the economists with respect to *the efficient breach of contract* theory and a strictly legal approach in respect thereof.

This critical overview is quite challenging and therefore it should benefit of more attention in the Romanian legal literature as well.

3 Methodology

Given the interdisciplinary approach of the topic, this paper will address both the economic and legal perspectives with respect to the impact of *the efficient breach of contract* mechanism, as developed by the economic analysis of contract doctrine. To this end, we will consider the following structure:

(i) *primo*, listing the main ideas and views of the specific mechanism developed by economic analysis of the contract doctrine (*i.e.*, *the efficient breach of contract* theory),

(ii) *secundo*, analyzing whether such ideas and views have a correspondent in the legal system of France and Romania, and

(iii) *tertio*, conducting an investigation concerning the different compatibilities and incompatibilities between the economic analysis of contract and the (strictly) legal approach of the matter.

This paper presents how *the efficient breach of contract* mechanism operates from a theoretical perspective, but also from a practical standpoint. The examples used here are intended to illustrate and make more comprehensible the application of this mechanism.

In light of the background described above, the presentation focuses on finding possible responses to the questions listed in the introductory section (Section 1), by analyzing:

- (i) the incidence of *the efficient breach of contract* theory,
- (ii) the incidence of the *favor contractus principle*, and

(iii) the possibility to reach a balance/equilibrium of the contract with the aim of avoiding the termination thereof or its onerous performance by one of the parties.

The approach of the presentation is transnational, as we refer to the legal system of Romania, France and United Kingdom.

4 Analysis

4.1 General considerations with respect to the economic analysis of the contract

The economic analysis of law studies the efficiency of the law strictly from an economic standpoint (Fabre-Magnan, 2010, p. 90). The economic analysis of contract is only a component thereof dealing with contract law. From a legal perspective, the main limitation/critique concerning the approach taken by the economic doctrine consists of the fact that the economic benefit to be obtained by the parties is the only purpose of a contractual relationship.

The economic analysis of the contract expressed views with respect to the contractual field from the very stage of the conclusion of the contract, continuing with the interpretation of the contract, with the performance thereof and ending with the termination of the contract, including the breach of the contract and the relevant damages to be paid in such a case.

For the purpose of this study we will further describe *the efficient breach of contract* mechanism.

4.2 For or against the legal recognition of the efficient breach of contract mechanism?

One of the most interesting and challenging theories/mechanisms developed by the economic analysis of the contract doctrine is the so-called "the efficient breach of contract". This mechanism pleads for the breach of the contract by one of the parties if at the moment of the effective performance thereof (by opposition to the situation envisioned as of the conclusion of the contract), the relevant party identifies a better way for placing the resources (which should otherwise be invested in the performance of the contract) and the profit obtained is higher than the costs related to the indemnification of the contractual partner from the ongoing contract - *i.e.*, the payment of the relevant damages to the other party for breach of the contract (Fabre-Magnan, 2010, p. 691-692).

Posner proposed the following example for illustrating the operation of this mechanism:

"Suppose I sign a contract to deliver 100,000 custom-ground widgets at 10¢ apiece to A for use in his boiler factory. After I have delivered 10,000, B comes to me, explains that he desperately needs 25,000 custom-ground widgets at once since otherwise he will be forced to close his pianola factory at great cost, and offer me 15¢ apiece for them. I sell him the widgets and as a result do not complete timely delivery to A, causing him to lose \$1,000 in profits. Having obtained an additional profit of \$1,250 on the sale to B, I am better off even after reimbursing A for his loss, and B is also better off. The breach is Pareto superior. True, had I refused to sell to B he could have gone to A and negotiated an additional step, with additional transaction costs – and high ones, because it would be a bilateral-monopoly negotiation. On the other hand, litigation costs would be reduced." (Posner, 2003, p. 120)

Similar to the case of the French jurists (Rochfeld, 2008, p. 118), discussing with Romanian jurists the meanings and implications of the *efficient breach of contract* theory risks raising difficulties with respect to understanding the concept. Moreover, such discussions could lead to contradictions and surprises in light of the classic *pacta sunt servanda* rule due to the fact that pursuant to the Romanian legal system (similar to the French system) the execution of the contract in kind (*i.e.*, the specific performance) represents the rule, while the equivalent performance by payment of compensatory damages constitutes the exception.

In addition, pursuant to the French doctrine (Rochfeld, 2008, p. 169), *the efficient breach of contract* risks to be incomprehensible for the French jurists also from a terminological standpoint as the consecrated expressions used with respect to the "destruction" of the contract make reference to "the termination for cause of the contract", "the termination for default of the contract", "the unilateral termination of the contract" etc. The same terminological issues apply *mutatis mutandis* for the Romanian legal system as well, as this topic benefit of a very poor reflection in the legal doctrine.

From the perspective of the features of this theory, the economic benefits resulted from the conclusion of the contract represent the sole purposes to be attaint by the contracting parties. Therefore, these aims are not satisfied anymore, the best option for the relevant contracting party is to fail to perform the contractual obligations undertaken. It results from the above cited example from Posner that this does not necessary equal to the termination of the contract (*i.e.*, it could only trigger, for example, a delay in the performance of the contractual obligations).

However, whereas the risk to terminate the contract as a consequence of the application of *the efficient breach of contract* mechanism (by opposition to the continuation thereof without obtaining the contemplated profit) is considerable, we will focus our below analysis on this hypothesis which involves several sensitive issues.

In light of the fact that the contract is mainly a vehicle for the economic exchanges, a termination thereof on grounds of lack of economic efficiency seams, *prima facie*, legitimate. Another argument supporting such an approach is represented by the fact that such termination would be accompanied by the payment of damages by the defaulting party. Therefore, even though the breach of the contract equals to the termination thereof, this is made in exchange of the payment of damages to the harmed party. In light of these short considerations, one could be persuaded that this theory is of great benefit from a legal standpoint as well.

But the question is: Does this perspective comply with the recent doctrinal developments mirrored in the French system and in the Romanian system which support the need for the maintenance of a certain contractual balance? *Prima facie* the answer is affirmative: the contract is not efficient anymore (at least) for one of the contracting parties and thus the balance has been compromised. The continuation of the performance of the contract is not stimulating for (at least) one party. Therefore, from an economic perspective, it can be argued that a so-called *efficient breach of contract* is benefic for both parties (Fluet, 2008, p. 156). On one hand, the party who

is breaching "efficiently" the contact is satisfied for escaping from a contract with the payment of damages and considering the latter more advantageous than performing the contract in the future. On the other hand, the other party is content for obtaining the damages rather than for having the contract further performed by the other party.

This preliminary conclusion based on a quick overview seems to make from *the efficient breach of contract* developed by the economists a very useful instrument to be applied by the jurists as well. However, albeit this preliminary conclusion appears seducing, a more in-depth analysis from a legal perspective risks to show important incompatibilities for recognizing and enforcing in practice this theory.

The main critique brought by the jurists to the economic analysis of contract (and of law, in general) concerns the individualist vision taken by such with respect to the contract. In this economic view, the contract represents only the vehicle aimed to assure the satisfaction of the egoist economic interests of the parties.

However, the recent French legal literature shows that the economists commence to consider the critiques expressed by the jurists, recognizing the limits of an individualist vision of the contract and, at the same time, acknowledging the "social individual" (Favereau, 2008, pp. 21-39).

The theory of *the efficient breach of contract* faces real obstacles in being accepted and implemented in the national legal systems setting forth the rule of the execution in kind of the contractual obligations (*specific performance*), while equivalent performance by means of payment of damages applies only in exceptional hypotheses (this being the case of the French (Jauffret-Spinosi, 2008, p. 25) and Romanian legal systems). To the contrary, in the common law system (in force in the United Kingdom) the situation is different as (i) the execution of the contractual obligations in kind (the specific performance) is the exception, the execution by the payment of damages being the rule (Jauffret-Spinosi, 2008, p. 19), (ii) *pacta sunt servanda* rule is in a certain manner "relative" (Jauffret-Spinosi, 2008, p. 19), and (iii) the contract is conceived rather as an instrument of patrimonial exchanges than of parties' consents.

In respect of the French law, the legal doctrine (Rochfeld, 2008, pp. 171-172) added to the above-mentioned ground (Rochfeld, 2008, p. 182) other several contraarguments against the application of *the efficient breach of contract* theory, as follows: (i) the impact of *pacta sunt servanda* principle, (ii) the moral grounds supporting the maintenance of a certain contractual justice; (iii) the safeguard of extra patrimonial elements; (iv) the difficulties related to the computation of the value of the damages to be awarded to the injured party; (v) the difficulties concerning the anticipation of the damages by the party contemplating the "efficient" termination of the contract; (vi) the absence from the French legal system of express provisions regulating two mechanisms deemed appropriate for the implementation of the theory: the creditor's obligation to find a substitution provision and the obligation of mitigation of damages etc.

Almost all the arguments listed above are applicable *mutatis mutandis* in the Romanian legal system as well. However, note should be made that by opposition to the French law, the New Romanian Civil Code (in force as of October 01, 2011)

expressly sets forth in Art. 1534 para. 2 the debtor's obligation to avoid and mitigate his damages.

The instruments of the Romanian law which seem to plead against the intentional termination of the contract, albeit such would be considered efficient (or not), are the provisions of Art. 1271 ("The parties are bound to perform their obligations even though the performance thereof became more onerous, either due to the increase of the costs corresponding to the performance of her own obligation or due to the decrease of the value of the prestation of the other party.") and Art. 1533 "The debtor is liable only for the losses that he has foreseen or that he should have foreseen as a consequence of the non-performance of the contract at the conclusion of the contract, except if the failure to perform is deliberate or is due to his gross negligence. Even in this latter case, the damages comprise only what it is the direct and necessary consequence of the non-performance of the obligation.") of the New Romanian Civil Code. In the French legal system, Art. 1150 ("The debtor is liable only for the damages that he has foreseen or that he should have foreseen at the conclusion of the contract, provided that the non-performance of the contract is not due to his intention.") of the Civil Code comprises similar provisions as those set forth in Art. 1533 of the Romanian Civil Code.

An analysis of the above mentioned legal provisions leads to the conclusion that the situation of intentional failure to perform the contract is to a certain extent covered by the Romanian and French Civil Codes. In light of theses provisions the reasons for which the debtor fails to deliberately execute the contract are irrelevant (provided that they do not concern *a force majeure* event, hypothesis which would be regulated by other legal provisions). Therefore, pursuant to the Romanian law, the sanction applicable in case of debtor's deliberate failure to execute the contract (grounded in the case at hand on the desire to renounce at the ongoing business relationship for another one deemed more stimulating) could be either (i) the debtor's opposition to the termination of the contract and the request to obtain the enforcement in kind of the contract (if possible), either (ii) the pecuniary sanction concretized in obtaining a higher value of the damages as in the case of non-performances which were not caused deliberately or as a consequence of acting with gross negligence (considering that the debtor is held liable for the prejudice which was unforeseeable at the conclusion of the contract).

Consequently, the prior estimation by the debtor of the damages to be paid by the latter risks to be quite difficult as the debtor could find himself in the impossibility to be aware of all the (unforeseen) losses incurred by the creditor. However, this estimation is the instrument helping the debtor to choose between (i) the "efficient" breach of the contract accompanied by the profit to be obtained from another contract to be concluded and (ii) the performance of the ongoing contract (Pop, 2009, p. 90).

In another respectable opinion, a part of the Romanian doctrine analyzing the compatibility of *the efficient breach of contract* theory in light of the Code civil of 1864 in force before October 01, 2011 (and particularly of the legal provisions concerning the performance of *facere* and *non facere* obligations by means of paying the relevant damages and the determination of the damages owed to the debtor as a consequence of the non-performance of its contractual obligations) reached the

conclusion that this theory was compatible with the Romanian legal system (Pop, 2009, pp. 89-91).

Irrespective of the pros and cons concerning the effectiveness of *the efficient breach of contract* in the Romanian and French legal systems, it is certain that the day to day business environment is familiarized with terminations of contractual relationships caused by the lack of financial efficiency thereof. These situations cover as well that type of contract termination lacking contractual basis (*i.e.*, unilateral termination of contract expressly allowed in certain circumstances) and therefore representing terminations for cause of the contract.

Consequently, the *efficient breach of contract* constitutes a mechanism used in the business environment, either if recognized or not *per se* in the applicable legal provisions from Romania or France.

The rigorous verification of the conditions of enforcement of such a mechanism is the only real control that can be exercised in respect thereof. For example, if the termination of the contract is judiciary (*i.e.*, submitted to the control of a court of law or arbitral tribunal), the judge/arbitrator will have to verify in a strict manner if the grounds justifying the termination are real and effective. From this perspective the party for which the contract become inefficient from an economic perspective might not have the legal instrument (as reflected in the applicable law) needed for the legal termination of the contract or might misuse of such right by trying to escape from a contract by invoking minor non-performances of the other contracting party. An illustrative case in this sense is offered by the French case-law in the decision rendered by Versailles Court of Appeals on June 8, 2000 - Versailles, June 08, 2000, SARL Sud est motors c/ SA Volvo automobiles France - (Rochfeld, 2008p. 180). In such decision the court qualified as abusive the termination of the contract by one of the parties on the basis of artificial grounds (by invoking the failure of a negotiation) animated by the sole purpose of choosing another contractor.

In any case, terminating the contract is the option which serves less a pleading in favor of the maintaining of a contractual balance during performance of the contract. Moreover, different legislations of European countries as well as model laws drafted at the European level with the aim of harmonizing the national legislations seem to allow the emanation of a *favor contractus principle* which pleads for the maintenance of the contract and thus making the termination thereof more difficult. *The favor contractus principle* plays when the performance of the contract is in a certain manner endangered. In such a case, the main issue is to establish whether the contract should be executed irrespective of the fact that the obligations incumbent on one of the parties are more onerous (and less satisfactory from an economic standpoint) or whether the contract should be terminated (with the payment of damages by the defaulting party – if the case).

Different European national legal systems as well as model laws developed at the European Union level for the purpose of harmonizing the different approaches of the national legislations set forth specific provisions which seem to be more favorable to the maintenance in force of the contract (albeit from an economic point of view the loss of the party bound to perform its obligations - become more onerous - is higher

than the loss which should be incurred as a consequence of the termination of the contract).

An effective analysis of the *favor contractus principle* will imply extracting from the above-mentioned model laws/legal systems the main provisions and/or principles (*i.e.*, hardship, reduction of prejudice etc.) which plead for the renegotiation/adaptation of the contract instead of the termination thereof.

In the spirit of the motto cited in the introduction of this study, a third option could be considered by the parties dealing with circumstances justifying *the efficient breach of contract*, such as identifying the means for "re-equilibrating/re-balancing the contract" for the purpose of avoiding the termination thereof or an extremely onerous performance of the contractual obligations by one of the parties.

5 Conclusions

It results from the above study that the reasons of *Homo Economicus* are not always congruent with those of *Homo Iuridicus*. *Homo Iuridicus* tries to overrun the economic egoist purposes related to the performance of a contract by inquiring himself about the necessity that each party considers the interests of the other party with respect to the contract. The research conducted in this respect lead to the conclusion that there are several incompatibilities and deficiencies of the approach of the economic analysis of the contract (and of law, in general).

In addition, several conclusions could be listed as valid both for the Romanian and the French legal systems:

- (i) The laws in force do not expressly regulate the efficient breach of contract.
- (ii) The execution in kind of the contractual obligations is the rule in the positive law. *The efficient breach of contract* mechanism is in contradiction with this view; from this standpoint there is a structural incompatibility between the application of this theory and the Romanian and French legal systems. Therefore, the creditor can make use of the means provided by the applicable law in order to ensure the performance in kind of the contractual obligations incumbent on the debtor. In case the preventive and palliative instruments aimed to assure the intended purpose (*i.e.*, the performance of the contract) fail, the (imperfect) sanctions set forth by the legislation are the computation of damages, the debtor being also liable for indemnifying the creditor for the loss which was unforeseeable at the conclusion of the contract. In such hypothesis the market rules and constraints will find other (indirect) "punitive" measures in order to proscribe the "incorrect" players, if the case.
- (iii) Law has also a moral component. This approach is not congruent with the failure to perform the obligations undertaken.
- (iv) The failures of the debtor to satisfy the legitimate expectations of the creditor could trigger our "compassion". On the other side, debtor's option for the

efficient breach of contract is a risky choice for the debtor as well, as the latter could be in the end in the situation to pay damages (to the debtor) having a higher value that the one initially estimated due to possible unforeseeable losses incurred by the creditor (but not taken into consideration by the debtor at the time of the decision to breach the contract).

- (v) Therefore, *the efficient breach of contract* involves several risks, advantages and disadvantages for all the parties involved.
- (vi) In certain cases this type of "destruction" of the contract could be beneficial for both parties, thus a certain precaution is indicated before proscribing the practical role of *the efficient breach of contract*.
- (vii) The parties could choose to protect themselves as of the conclusion of the contract of possible undesirable consequences triggered by a contemplated *efficient breach of contract* (a) by setting forth in a very clear manner the cases allowing the termination of the contract and the penalties applicable in such a case (which must be as high as possible in order to be persuasive) or (ii) by providing certain mechanism/techniques aimed to postpone the definitive effects of the contract/conclusion of the contract (Kerhuel, 2010, p. 360). All these possible clauses could be drafted either for the purpose of assuring a facile way to escape from the contract, or, to the contrary, for the purpose to make the termination thereof more difficult.
- (viii) The contractual balance is rather maintained by adapting the contract than by "destroying" it.

References

Cathiard, A. (2006) L'abus dans les contrats conclus entre professionnels: L'apport de l'analyse économique du contrat, Presses Universitaires D'Aix-Marseille, Aix-en-Provence

Fabre-Magnan, M. (2010) Droit des obligations. 1 – Contrat et engagement unilatéral, second edition, Presses Universitaires de France, Paris

Favereau, O. (2008) Qu'est-ce qu'un contrat? La difficile réponse de l'économie, in Jamin, Ch. (ed.), *Droit et économie des contrats*, Librairie générale de droit et de jurisprudence, Paris, p. 21-42

Fluet, C. (2008) La rupture efficace du contrat, in Jamin, Ch. (ed.), *Droit et économie des contrats*, Librairie générale de droit et de jurisprudence, Paris, p. 155-167

Jauffret-Spinosi, C. (2008) Les grands systèmes contractuels européens, in *Le contrat en Europe aujourd'hui et demain,* Société de Législation Comparée, Paris, p. 9-28

Kornhauser, L., Du Marais, M., Selmouni, K., Lanneau R. (2010) L'analyse économique du droit: fondements juridiques de l'analyse économique du droit, M. Houdiard, Paris

Lando, O. (2008) The contract in Europe: today and tomorrow. Final remarks, in Cabrillac, R., Mazeaud D., Prüm, A. (eds.), *Le contrat en Europe aujourd'hui et demain*, Société de Législation Comparée, Paris, p. 177-191

Mackaay, E. (2000) L'analyse économique du droit, Vol. 1 Fondements, Bruylant - Thémis, Bruxelles - Montréal

Mackaay, E., Rousseau, S. (2008) Analyse économique du droit, second edition, Dalloz-Thémis, Paris-Montreal

Pop, L. (2009) *Tratat de drept civil. Obligațiile, Vol. II. Contractul*, Universul Juridic, Bucharest

Posner, R. (2003) *Economic analysis of law, sixth edition*, Aspen Publishers, New York

Rochfeld, J. (2008) La rupture efficace, in Jamin, Ch. (ed.), *Droit et économie des contrats*, Librairie générale de droit et de jurisprudence, Paris, p. 169-192

Vasilescu, P. (2003) Relativitatea actului juridic civil. Repere pentru o nouă teorie generală a actului de drept privat, Rosetti, Bucharest

Vasilescu, P. (2007) Actul juridic civil, in Reghini, I., Diaconescu, S., Vasilescu, P. (eds.), *Introducere în dreptul civil, second edition*, Sfera Juridică, Cluj-Napoca

Vasilescu, P. (2012) Drept civil. Obligații, Hamangiu, Bucharest.

Introduction to the study of public transport accessibility

József Pál Lieszkovszky

PhD student, Doctoral School for Regional and Economic Sciences, Széchenyi István University, Győr, Hungary Email: lieszko@gmail.com

Abstract. From the aspect of regional development or regional planning it is very important to analyze of public transport accessibility (rail and road passenger transport services equally). Based on this research should be evaluated the quality of public transport towards central locations. In this paper we gain insight briefly into the methodology of public transport accessibility then it will be presented the bus availability conditions among the Micro Region Centres (LAU1 level) and Micro Region Settlements (LAU2 level) of selected sample area (Western Transdanubian Region of Hungary – NUTS2 level) by the support of availability indicator (Index of Zimpel).

Keywords: Hungary, accessibility, public transport, transport geography

1 Introduction

The most important spatial functions of public transport availability are the "transport connection" and the "transport attachment".

- The transport connection means the mutual public transport possibilities between the same size of settlements (usually cities) which have intra or inter-regional aspects. It is important to emphasize the mutual indicator because the connection takes place not only from one direction but both sides involved. The transport connection is a horizontal type of connections.
- The transport attachment means the hierarchical assignment of smaller municipalities to the higher centres by the help of the transport system. The transport attachment is a vertical type of connections and could be manifested at the catchment area.

Substantially, the urban-rural partnership is not able to exist without any transport possibilities therefore transport has a crucial role in the establishment of catchment areas. In addition, the mobility of local population of areas beyond the catchment areas will increase if they are involved into the city-center transport system. Thus, the transport attachment can cause the concentration of activities or even equable distribution in the catchment area (*Erdősi* 1985, *Næss* 2006).

2 Literature review: The functions of spatial accessibility

The delimitation of catchment areas is on the focal point of the Hungarian researchers from the '50s. Two important Hungarian settlement geographers' studies are worth mentioning: Tibor Mendöl¹ and Pál Beluszky². Mendöl said that the ratio between the size of city and catchment areas depend on the development of transport. Therefore that area is called catchment area where the city "does" or "has" the central functions, not the separate settlements (*Mendöl* 1963). However, Beluszky called the transport catchment area the potential catchment area, from where the central settlements can be reached by the shortest time and the most frequent and cheapest way in spite of other central locations (*Beluszky* 1974).

Investigating the catchment area by the aspect of two main components of passenger transport, - rail and road passenger transport services - we can conclude that the urban centres of Hungary - apart from the urban centre of capital city - are not large enough to have a fast rail connection with their suburb/catchment areas. Contrarily, the organization of bus transport network has far more catchment aspects than the rail passenger transport network and the smaller cities are also the centres of those bus lines which operate within the catchment area (*Erdősi* 2000). In fact, however, the smaller the centre's catchment area, the more likely to have connections with the catchment area's settlements by the interregional bus lines only and the ratio of short route (catchment) transport is negligible.

3 Data and methodology of public transport accessibility

The definition of inter-municipal public transport system can be found in several studies of transport geography. This measurement can be done by different methods: using hierarchical system of central places by Christaller (*Bromley–Bromley* 1979, *Preston–Rajé* 2007), using the gravity model (*Simon–Csatári* 1982) or determining the size of catchment areas by time zones (*Szörényiné* 1985).

¹Tibor Mendöl (1905–1966): One of the prominent researchers of the Hungarian Science of Geography. His work is currently standard at the field of settlement geography.

²Pál Beluszky (1936–): He is the most known alive settlement geographer in Hungary. His research area covers the villages, catchment areas and human geography.

The transport accessibility has three important components: the distance, the travel time and the frequency of public transport vehicles. The travel time and frequency as two important integral components of complex public transport accessibility indicator were applied first time by Zimpel (Erdősi 2000) who tried to determine the public transport structure of Bavaria with this method.

Index of Zimpel

$$V = F_m \frac{1}{f_m}$$

where

V = the transport value of investigated region's settlement Fm = the average time of access the centre by public transport (in minutes) fm = average daily frequency (number of buses on workdays)

Hungarian researchers have already used similar indicators like the index of Zimpel (Erdősi-Horváth-Kovács 1986, Szörényiné 1989). Mrs. Szörényi also used an index quite similar to Zimpel to determine the accessibility of rural areas at three Western Hungarian counties (Győr, Vas, Veszprém). The accessibility of three administratively controlled functions (council seat, centre of primary schools and centre of medical facilities) was investigated in that research at the above-mentioned area. Depending on the presence of these functions represented in that settlement, the centers were categorized into three different types. Accessibility is measured by the ratio of the average travel time and the average daily frequency (number of buses on workdays). As Zimpel thought, the average daily frequency is in direct proportion to the quality of accessibility and the travel time is in inverse proportion to the quality of accessibility therefore it can be used as appropriate metrics. The main conclusion is that the extension of settlement relationships does not depend on the numbers of central functions, in other words the allocated functions of small centres do not determine their accessibility but the accessibility rather depends on the proximity of job centres or larger cities as well as the location of settlements on main road networks (Szörényiné 1989).

There are more other calculation methods for public transport accessibility of settlements which approach the concept of accessibility from different perspectives. There is a research which deals with the transport opportunities of residents of southern Hungarian villages (in Baranya County) which focuses on the accessibility of local government and medical centers. (Buses between 5am and 10pm were taken into account.) The average "waiting time" was determined to compare the public transport connections of Baranya County's settlements (in the relations above) and

this "waiting time"³ is part of the whole transport time. The short conclusion is if the waiting time is high it means that the number of buses is low or erratic therefore it needs waiting for the bus very long time at some parts of the day, mainly in daytime. There were defined 10 different groups based on the waiting time and the neighbourhoods were classified by this method (*Erdősi–Horváth–Kovács* 1986).

Zimpel's assumption is that within the catchment area the intensity of links of public transport decreases away from the central settlement. In essence, this approach is similar to the theory of concentric circles (*Lengyel-Rechnitzer*, 2004). This approach is used when investigating the study area.

The sources of datas are from the Hungarian online timetable system (www.menetrendek.hu) and they were collected in November 2011.

4 Study area: The Western Transdanubian Region of Hungary

Hungary has seven NUTS2 Regions, one of them is the Western Transdanubian region which is located at the northwest part of the country. These regions consist of three NUTS3 level areas: Győr-Moson-Sopron County, Vas County and Zala County⁴.

It will be presented the size and the distribution of direct bus contacts between Micro Region Centres and Micro Region settlements⁵ in this part of the paper⁶. Four categories were defined: the lighter the colour of settlement is on the map, the better the Index of Zimpel is, which means that high bus service numbers are associated with short access time. The darker the colour is, the indicator of Zimpel Index is getting worse. In that case there are associated decreasing numbers of bus services with increasing access time. Those settlements are marked grey from which you need a transfer to get to the Centre of Micro Region. (The connection time was maximized in 20 minutes). Settlements are marked with black have no direct bus connection with the Centre of Micro Region or you need a multiple transfer. Settlements marked with circles are the Centres of Micro Regions.

³ "Waiting times": how many minutes do I have to wait for a bus after finishing my activities in the central settlement and I want to go back to my home village

⁴ Details in Appendix

⁵ Settlement means local administrative unit

⁶ Accessibility of railway was also examined but it was not depicted on maps. It is mentioned at the oral presentation.

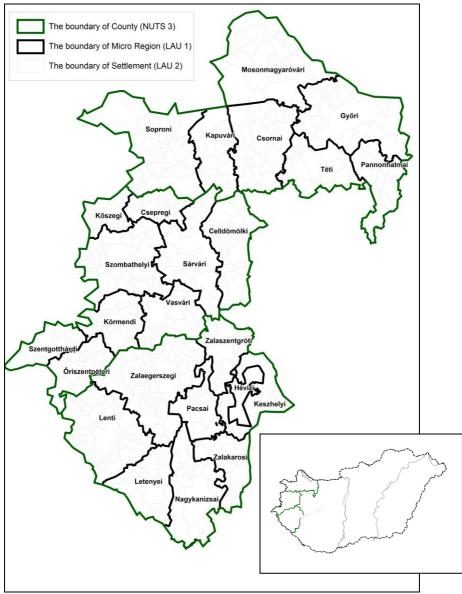
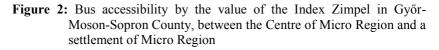


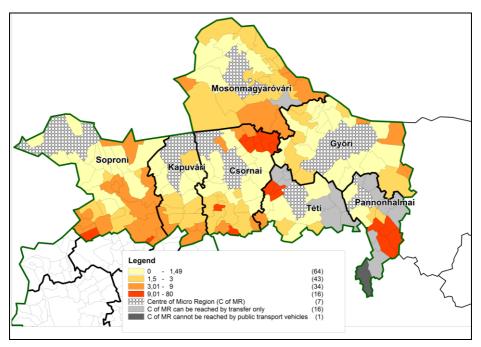
Figure 1: The map of Western Transdanubian Region and Hungary

Source: edited by the Author

4.1 Bus accessibility in Győr-Moson-Sopron County

It is obvious that nearly all of the settlements of Győr Micro Region (where the county seat of Győr-Moson-Sopron County is located) have strong connection with the County seat, Győr. In contrast, Pannonhalma and Tét Micro Regions are located southern to Győr Micro Region. They are typical instances of improper drawing of border lines of Micro Regions. Both settlements of those Micro Regions have much better connection to the County Seat (Győr) than to their Centres of Micro Regions. (City of Pannonhalma can be reached mostly by train from the settlements of its Micro Region). In the Sopron and Kapuvár Micro Regions the Zimpel Index has expected values as we are getting away from the Centre of Micro Region. (The map shows the settlements located in transport shadow which can be seen clearly at Sopron Micro Region). At the north-eastern corner of Csorna Micro Region there are situated some settlements which have very bad values in spite of being closer to the Centre of Micro Region. It means that the access time is short and the numbers of bus services are very low. However their connections to Győr are much stronger.



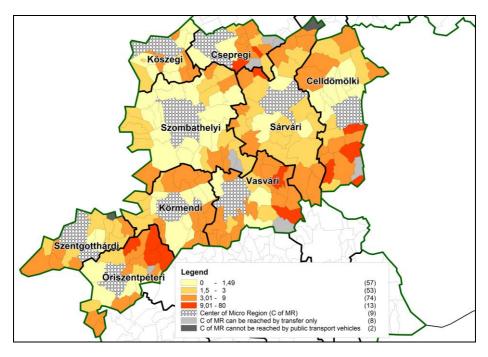


Source: edited by the Author

4.2 Bus accessibility in Vas County

The different areas like concentric circles become distinct at the Szombathely Micro Region away from the County Seat, Szombathely. We can find the same situation at the Sárvár Micro Region. The settlements of Csepreg Micro Region have very hectic public transport links to the Centre of Micro Region. (Train does not run here). We can see some interesting areas that need comment: The settlements located at the south of Celldömölk Micro Region have very high (bad) value of Index Zimpel because they have bad bus connection with the Centre of Micro Region but also these settlements are situated near major railway lines so the train is the main public transport vehicle (towards the Centre of Micro Region). Real remote areas can be found at the northeast part of Őriszentpéter Micro Region. (In the Micro Region of Szentgotthárd there is a very tiny settlement which has no public transport connection to any other settlements, it is marked grey.)

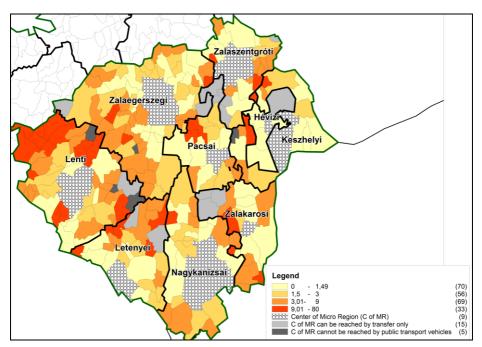
Figure 3: Bus accessibility by the value of the Index Zimpel in Vas County, between Centre of Micro Region and a settlement of Micro Region



Source: edited by the Author

Glancing at the map, the first thing that we notice is the red continuous settlement area at the north side of Lenti Micro Region. These tiny settlements have just 1-3 daily bus connections with Centre of Micro Region (Lenti) so they create a public transport "ghetto" area. (You can find here the highest value of Index Zimpel - 80 - at Magyarföld, which means that there is only one direct bus connection with the Centre of Micro Region, and the duration of this route is 80 minutes). Moreover, in the Lenti Micro Region there are three other settlements which have no connection with the Centre of Micro Region. In 2007 there were created some new Micro Region (Pacsa, Zalakaros) but due to the new boundaries, some public transport accessibility problems occurred. These new LAU1 levels were "put together" by mostly political volition; some of them – like the Micro Region Hévíz – has some special situations. (Settlement of Rezi is adjacent to City of Keszthely but belonged to the Hévíz Micro Region, however City of Hévíz can be reached by transfer at City of Keszthely)

Figure 4: Bus accessibility by the value of the Index Zimpel in Zala County, between Centre of Micro Region and a settlement of Micro Region



Source: edited by the Author

5 Conclusions

We can conclude that theory of Zimpel prevails mainly at the Micro Regions of County seats (Győr, Szombathely) but can be seen at smaller Micro Regions either (Kapuvár Micro Region). Some settlements have no connections with their Centre of Micro Regions due to the improper drawing of border lines of Micro Regions. (This rate is 25 % in the Pannonhalma and Tét Micro Regions). These problems were not able to be solved at the newly created Micro Regions moreover these new Micro Regions have huge proportion of settlements which need transfer to reach the Centre of Micro Region. In addition, although it is not widespread, here are some public transport exclusion areas – called public transport "ghetto" areas – but those settlements do not belong to the group of settlements excluded socio-economically.

6 References

Beluszky, P (1974): Nyíregyháza vonzáskörzete. A város-falu közötti kapcsolatok jellege és mennyiségi jellemzői Szabolcs-Szatmár megyében [Agglomeration of Nyíregyháza. Quantitative and qualitative characteristics of urban-rural relations in Szabolcs-Szatmár County]. *Földrajzi Tanulmányok [Geographical Studies]*, 13. 118-150. p.

Bromley, R – Bromley, R.D.F (1979): Defining Central Place Systems Through the Analysis of Bus Services: The Case of Ecuador. *The Geographical Journal*, Vol. 145, No. 3. 416-436. p

Erdősi, F (1985): Közlekedés és vonzáskörzetek [Transport and catchment areas]. In Faragó–Hrubi (szerk.): *A vonzáskörzetek gazdasági és közigazgatási kérdései*. [In Faragó–Hrubi (eds): *The economc and administrative issues of catchment areas*] MTA Regionális Kutatások Központja Dunántúli Tudományos Intézete, Pécs [MTA HAS Transdanubian Research Institute, Pécs] 223–232. p.

Erdősi, F – Horváth, Cs – Kovács, K (1986): A községekben élők közlekedési lehetőségei Baranya megyében [Transport opportunities of residents in villages of Baranya County] – *Területi Statisztika [Spatial Statistics]*, 36. 1986. 3. 220–240. p.

Erdősi, F (2000): A kommunikáció (közlekedés-távközlés) szerepe a terület- és településfejlődésben. [The role of communication (transport-telecommunications) at regional and urban development] VÁTI, Budapest

Gray, David – Shaw, John –Farrington, John (2006): Community transport, social capital and social exclusion in rural areas. *Area.* Vol 38.1, 89–98 p.

Lengyel, I – Rechnitzer, J (2004): *Regionális gazdaságtan. [Regional Economics]* Dialóg Campus, Budapest-Pécs

Mendöl Tibor (1963): *Általános településföldrajz* [General Settlement Geography]. Akadémiai Kiadó Budapest

Næss, Peter (2006): Accessibility, Activity Participation and Location of Activities: Exploring the Links between Residential Location and Travel Behaviour. *Urban Studies*. Vol 43. No 3, 627-652 p.

Simon, I – Csatári, B (1982): A települések közúti és vasúti kapcsolatainak földrajzi vizsgálata egy forgalmi árnyékban lévő területen. [The investigation of rail and vehicular connnections of settlements at a transport shade area]. *Alföldi tanulmányok* [*Great Plain Studies*]. Vol 6, 175-194 p.

Szörényiné Kukorelli, I (1985): A közlekedésföldrajzi gravitációs zónák meghatározása és értékelése az Észak-Dunántúlon. [The identification and evalution of transport geographical gravity zones at Northern Transdanubia]. *Földrajzi Értesítő. [Geographical Bulletin]* Vol 34/3. 325-358. p.

Szörényiné Kukorelli, I (1989): A falusi térségek elérhetőségi vizsgálata [The analysis of accessibility in rural regions]. *Győri Tanulmányok [Studies of Győr]*, 10. 1989/1. 81–96. p.

7 Appendix

NUTS 3	LAU 1	Population	Number of
(County)	(Micro Region)	(2010)	Settlements
Győr	Csornai	34.167	34
	Győri	182.321	27
	Kapuvári	23.395	18
	Mosonmagyaróvári	74.588	27
	Pannonhalmi	16.900	18
	Soproni	98.156	40
	Téti	18.908	19
Vas	Celldömölki	24.583	28
	Csepregi	10.736	16
	Körmendi	21.227	28
	Kőszegi	18.346	15
	Őriszentpéteri	6.549	22
	Sárvári	35.922	32
	Szentgotthárdi	14.611	15
	Szombathelyi	113.526	40
	Vasvári	13.864	23
Zala	Hévízi	12.781	8
	Keszthelyi	34.971	16
	Lenti	21.338	51
	Letenyei	16.645	27
	Nagykanizsai	65.647	27
	Pacsai	10.469	20
	Zalaegerszegi	96.368	65
	Zalakarosi	12.923	19
	Zalaszentgróti	17.449	20

 Table 1: Some datas of the Western Transdanubian Region

Source: edited by the Author

Author Index

Çiço, Betim, 588

Acquah, Sylvia, 274 Albaqawi, Hamdan, 704 Alharbi, Badr, 430 Alsaqri, Salman, 694 Alshammari, Farhan, 699 Andreea Bragă, Andreea, 689 Antovski, Ljupcho, 635 Atanasiu, Adrian, 625

Bačová, Viera, 764 Balan, Andreea Teodora, 326 Barkoukis, Vassilis, 772 Beck, Stephen, 784 Bekaj, Armend, 850 Belhorma, Souad, 685 Brânda, Oana Elena, 948 Bratanis, Konstantinos, 497 Buhaş, Raluca Sînziana, 725

Camelia, Straia, 261 Caradaică, Mihail, 923 Chipea, Floare, 725 Chirita, Cristina, 304 Christodoulou, Elias, 598 Christofis, Nikos, 189 Cowling, Anthony J., 453 Cowling, Tony, 551

Dautov, Rustem, 539 Dimitriou, Stylianos, 492, 598 DOTCHE, Koffi A., 645 Dranidis, Dimitris, 497 Drobná, Eva, 764

Eleftherakis, George, 453 Emil, Iudean Ovidiu, 858

Febriansyah, Hary, 253

Georgescu, Adela, 625 Gjergji, Elda, 670 Gonidis, Fotis, 565 Gourgoura, Esida Gila, 202 Gronostaj, Aleksandra, 793 Gusev, Marjan, 577

Hajra, Arben, 517 Harris, Peter, 772

Ismyrlis, Vasileios, 800 Istrefi, Dashmir, 588

Jones, Christopher, 784 Jovanovik, Milos, 504

Kakaevski, Ğorgi, 528 Kalyva, Efrosini, 735 Kamsaris, Dimitrios, 19 Kapetaniou, Chrystalla, 817 Karvounis, Alexandros, 77 Kechagia, Polyxeni, 62 Kefalas, Petros, 551 Khodzhibaeva, Mukhayo, 125 Kollias, Christos, 189 Korbi, Florenca, 935 Kostoska, Magdalena, 577 Kostovski, Martin, 504 Koturovic, Darja, 896 Koumparoulis, Dimitrios, 19 Kourtesis, Dimitrios, 539, 565 Kriel, Elli, 963

Ladas, Aristea I., 747 Latif, Atif, 517 Lazouras, Lambros, 817 Lazuras, Lambros, 772 Lieszkovszky, József Pál, 1025 Ljepava, Danimir, 375

Maćków, Weronika, 980 Machek, Ondrej, 208 Magyar, Maria, 343 Marti, John, 388 Meier, Petra S., 735 Memushi, Armando, 91 Metaxas, Theodore, 62 Mišev, Anastas, 528 Mihaiu, Simona Ionela, 657 Milić, Boja, 318 Milošević, Lana, 228 Monostori, Zoltán, 167 Moreno, Flor, 45 Mynařková, Lenka, 710

Nádvorníková, Pavla, 823 Nikolaidou, Eftychia, 125 Nikolic, Ivana, 213

Olimid, Ruxandra, 610 Onojescu, Alexandru, 873

Papadamou, Stephanos, 189 Papamalis, Fivos E., 735 Paraskakis, Iraklis, 539, 565 Petreska, Isidora, 475 Pireva, Krenare, 551 Polovina, Ana, 120 Popovic, Ljiljana, 120 Psychogios, Alexandros, 304, 326 Pulaj, Enida, 422 Pustina, Gert, 353

Radevski, Vladimir, 517 Rajtukova, Viktoria, 481 Ribarski, Pance, 635 Ridwan, Mohammad Sihab, 388 Ristov, Sasko, 577 Ropan, Delia-Ionela, 1013 Rousis, Konstantinos, 453 Rrokaj, Shezai, 987

Sadeghi, Maryam, 916 Salminen, Joni, 438 Sarmány-Schuller, Ivan, 730 Sava, Cecilia Alina, 888 Savva, Christos, 817 Serafini, Giovanni Oscar, 279 Shah, Noaman, 138 Shoraj, Dritan, 407 Shtembari, Eriona, 353 Siegal, Michael, 747 Simons, Anthony J.H., 497 Solomos, Dionysios, 19 Soteriades, Elpidoforos, 817 Soucek, Michael, 105 Stagiannis, Apostolos, 189 Sula, Adelina, 987

Túnyiová, Mária, 730 Tojtovska, Biljana, 464 Topçu, Güneş, 153 Toth, Teodor, 481 Trajanov, Dimitar, 504 Triantafylli, Danai, 776

Vasilaki, Athina, 326 Vivas, Ana B., 747

Williams, J.P.R., 784

Xhaferaj, Anjeza, 903

Zafirov, Goran, 177 Zela, Granit, 831 Zenelaj, Dorina, 407 Zenuni, Jaup, 994 Zivcak, Jozef, 481

E-Papers

The Case of Kosovo and Albanian Media from 1981 to 1999

Aferdita Sokoli

Albanian Media's Center of Cultural Value's Promotion. <u>aferditasokoli@yahoo.com</u>

Abstract: This study focuses on the coverage of the Media in Albania of the historical development in Kosovo from 1981 to 1999 and the role played by this media in the information of its public opinion. The Kosovo's problematic in Albanian Media from the beginning of the Spring1981 to the end of the Spring1999 is placed in different grounds and experiences of different social systems. The evaluation of its coverage during the years 1981 - 1990 takes into consideration the point of view of a sole media in conditions of the existence of the monist system, meanwhile the evaluation for the period 1990 - 1999 supports the perspective of the pluralistic media in the country.

The main findings presented in this research are: Self-determination remained a crucial issue in monist and even in the pluralist media. The war in Kosovo gave to the Albanian media new dimension. For the first time it faced to the challenges of media's war and gained the first experience of media's war.

Keywords: Media, reporting, constitution, autonomy, nationalism.

Introduction

The aim of this study is to expose the treatment's level of the Albanian Media regarding to Kosovo's issue during this period. In the same time it analyzes and evaluates the level to the media professional standarts. The study explain the main factors and the basic reasons have been affected on the that media's level each time in the study's period. It intent to point the physiognomy of agenda setting and media's sources as important aspects of the theoretical field of media's studies. It seeks to understand the reporting's ways of the Albanian media and it role in creating public opinion.

This study is divided into three main chapters.

The first chapter analyses the Kosovo problematic in the media of 1981 - 1999 and covers two main periods of the historical development of Albanian's media: communist and pluralist societies. It takes as well into consideration the qualitative changes of the media in this country and creating of the pluralist media as result of the fall of monist system and the radical changes in the Balkans region and Eastern Europe.

The most rational way to judges for the professional level of the engagement of media in this direction was the focus on its priorities. This process on its part was determined by the most vital of the Albanians in Kosovo in different phases of their historical development in the two decades in study. The issues of the rights for self-determination of Albanians in Kosovo and the ways of solving their current problem, as essential for the Kosovo problematic were in the focus of the Albanian Media. Respecting the chronological order of the events, the study in its first chapter analyses the coverage that media did to the process of cancellation of the constitutional status of the autonomous region during the period 1981-1990 and the alternatives of solving the Kosovo case from 1990 to 1999.

The second chapter describes the Kosovo Case as an experience in the theoretical and practical field for the Media of Albania. Regarding the problematic of years 1981-1990, Kosovo's agenda setting was decided by

the high level party instances, since media was entirely controlled by the party and there was a lack of editorial independence and alternative media. The pluralistic press gave new dimensions to the agenda setting and media's sources of Kosovo case.

The third part describes the role of the Media in Albania in creating of public opinion regarding to the Kosovo's issue, particularly during the war in Kosovo. In this relatively long period, important steps were undertaken in two directions: that of professional improvement and that of the consolidation of the Albanian media character. For the first time Albanian media treated a new aspect, that of the media in times of war that was unknown before.

1. "The Kosovo problematic in Albanian media during 1981-1999"

1.1 The process of gradual elimination of the status of the autonomous region of Kosovo and how it was reported in the media of the communist period in Albania (1981-1990)

The unprecedented dramatic developments in Kosovo became an important object of the information means in Albania. Even though delayed in broadcasting of information for this period, the Albanian Media started to treat these events as signals of the beginning of a regressive process in Kosovo with severe consequences for the existence of the Socialist Federative Republic of Yugoslavia.

The issue of protection and respecting of the principle of the right for self-determination of the Albanian population in Kosovo focused on the respecting the international law, in the historic circumstances of the existence of the Cold War. This universal right in the case of Kosovo took a special emphasis because "all the Albanian national movement is supported in the rights for self-determination"¹ and that request of Albanian was "in the spirit of the Constitution of Yugoslavia."² The refusal of the Albanian rights for self-determination in the Yugoslav constitutional right have been considered an axiom for the three Yugoslav constitutions of post Second World War. The media of Tirana this constitutional constant for the Albanian in Kosovo specified as "legal habit" of the federative constitutionalism, one "habit that excludes its responsibility for everything that is decided regarding with it in different phases and instances."³

In this media were evidenced basic reasons of the rejection of the Albanian request for the status of the republic. There were economic interests of Serbia and of the Federation of the Kosovo's natural resources, which vital in the circumstances of deep crises in Yugoslavia that time.

The other essential reason considered by this media related with the security interests that defined the Kosovo area "second echelon" of the strategic defense of Serbia "geostrategic center of the Balkans....a territory from which could undertake independent operations in Kosovo and from Kosovo."⁴

In the Albanian media took place the polemics with the Yugoslav media referring to different point of views regarding to the position of Albanians in Yugoslavia. Serbian media's interpretation proclaimed that Albanian population enjoyed the all rights as other citizens of the federation.

The opposite arguments set in the Albanian media. It underlined that although the spaces that the Constitution 1974 created and that "through some of its articles it ensured Kosovo the status of the republic, the reality was different. Kosovo was not a republic. With this reality the Albanians were faced every day."⁵ In its core this constitution contains inequality and the limitation for Albanians, the majority in Kosovo and the third population in the federation considering the number.

This status did not fulfill the request of the Serb minority in Kosovo, too and did not neutralize neither that of the Serb republic leadership, expressed in 1975 in the Blue Book. In this way, "during the autonomous period, neither the Albanians were satisfied, nor the Serbs. The Albanians aimed more, their political

¹ Lalaj A 1994 "E drejta per vetevendosje dhe shqiptaret e Kosoves" Studime Historike Tirane N 3 - 4

² Korabi A 1981 "Kerkesa Kosova Republike nuk mund te shuhet as me dhune, as me teorizime boshe" Zeri i Popullit Vol Korrik - Shtator N 218 (10331)

³ Ibid

⁴ Janjie D.J 1993 "National Identity, movements and serbian and albanians nationalism" "Interethnic Relations in Serbia/Yugoslavia: Alternatives for the Future" CEU Budapest p. 122.

⁵ ASHSH 2009 "Historia e Popullit Shqiptar" V 4 Toena f 389

requests were clear, while Serbs were afraid that now had lost the Kosovo and saw the last chance for regaining it."6

In the coherent stance toward the right of the Albanians for self-determination as "a request that would not fall by the end of the day until it takes the most right solution,"⁷ that "if it is not in the letter, it is in the spirit of the Constitution of Yugoslavia"⁸ in the analysis and in the media's prognosis in the very beginning it was expressed the doubt that this foundational document was risked by the Serb nationalist clan. Not without purpose in the Albanian editorials highlighted "if they seriously take this Constitution."⁹ This accurate prediction of the Albanian media after the demonstrations of Year 1981 was treated as a regressive process, in whose in its phases would concretize the strategy of the elimination of the autonomy of the region, through escalation of violation.

The Memorandum of the Scientific Academy of Serbia "a systematic and lucid summation of the various currents of Serb nationalist resentment"¹⁰ accelerated this process finalized in a bloody way by the decision of the Federative Assembly in January 1990.

Regardless of the treatment with objectivity of the situation in this media, the attention draws to the application of the methodology of intimation for the events, likewise were the northern Irish, Palestine's, German Reunification events, damaged the seriousness of the media's engagement. This phenomenon is explained with the fact of existence of a sole media, depended by the official politics which for the sake of the interstate relationships with the Yugoslav Federation limited the freedom of media's opinion by imposing the functioning of propaganda.

On the other hand, the treatment of the Kosovo case from this media served to the state party in order to develop the its ideological war against Yugoslav revisionism in protection of the Marxist principles for national case. This reality impacted a lot in narrowing of the dimensions of objectivity of treatment of the mentioned thematic.

1.2. Two alternatives of solving the Kosovo case : Peaceful manner and that of the armed war and the Albanian mediatic space (1991-1999)

This period analyses the printed and the TV broadcasting of the events with the emerging of the Albanian opposition Kosovo Democratic League, (KDL) its pacific way of solving the Albanian case and the other alternative of solving the case through the armed war (KLA –UCK) supported by NATO operations.

The creation of an Albanian opposition alternative in moments when the process of full elimination of the status of the autonomy was ending, oriented the monist media toward the activity of this alternative, leaded by the KDL. This event and the developments in general are slightly reflected, in the level of the news without the deep media's analysis, while the evaluation for this political factor was fragmented. In fact its political platform "was not published anytime in its generality"¹¹ and the media did not request for more and its perception about it was created through the press release and conferences of this political force.

Such media's reaction was conditioned by the priorities of the politics which evaluated as emergent the domestic situation in the country. Albania was facing social-economic difficulties, impossible to be solved within its social system, meanwhile that the great changes in Europe were also reflected in the country. The complexity of this situation had misplaced somehow the Kosovo agenda of the Albanian media. On the other hand it should be highlighted the fact that the Albanian opposition in Kosovo was anti-communist and the indifference of the communist regime was reflected even in the monist media.

With the issuing of the first opposite newspaper in Albania the treatment of the cases of Kosovo in media radically changed. The opposite media with its newspaper urged the peaceful resolution of the Kosovo case. The pacific way for solving Albanian issues on Kosovo proclaimed by leader Rugova was presented

⁶ Kraja M. 1995 "Vite te Humbura" Eurorilindja Tirane f 113

⁷ Korabi A. 1981 "Kerkesa Kosova republike nuk mund te shuhet as me dhune as me teorizime boshe" Zeri i Popullit Vol Korrik - Shtator N 218 (10331)

⁸ Ibid

⁹ Ibid

¹⁰ Benson L.B 2001 "Yugoslavia : A Concise History" Palgrave NY p. 146

¹¹ Kraja M.K 1995 "Vitet e humbura" Eurorilindja Tirane f 167

in Albanian Media through his statements in different meetings, press conferences and so on during 1991 – 1993. But there is no editorials or argumentative analyses which could explained widely and deeply the reasons of this Rugova's choice and if it was right way for the Albanian issue in former Yugoslavia. This media published only the Rugova's argument regarding to this question. Anyway in its beginning this media was very attentive at balance preserving. Regarding to the pacific way for solving the Kosovo question in this media was presented in the same shape the alternative opinion which considered as "a great illusion to be believed that with politics of cocktails could be created the state and to be accurately solved the national case."¹² Since 1992 year the protocol information for the KDL increased its dimension in this media by changing the relation with the deep analysis for this political force. By viewing the press of that time could say this phenomenon defined almost Albanian Media.

From 1993 and especially in 1998 in Kosovo coexisted two ways of solving the same case and two different political forces of the Albanian population were faced for the same aim. They were reflected in different ways in different media. "The war in itself was the result of the fact that the peaceful politics was exhausted. We could not achieve a pact with the international community that would stop Serbia and its actions came as an escalation toward one population. As a result, began a very legitimate reaction when the peaceful politics was exhausted"¹³ have been interpreted this phenomenon in Albanian Media by the Kosovo's political leaders. The idea of armed war, as a sole unused alternative of the Albanians against S. Milosevic in power was widely articulated in this media.

The role of the KLA reduced the role and the dimension of Rugova in Kosovo and as a result in the media. There was a media's clashes about him, particularly after his meeting with Miloshevic. The political communication among two forces was reflected even in the media of Tirana and in particular in that party media. The problematic in the respective political class from the two sides of the boundary interfered in political "supporters" in Albania or "dominance" of influence of the Albanian political specter in the political factor in Kosovo by creating in the eye of the international public opinion a false and ridiculous situation. Meanwhile titles of some notice of parties newspapers and articles where the KLA was accused as a "terrorist", "serbian agents" confused the public opinion. On the other hand, the "visit of Rugova in Beograd was without any doubts one favorite topic to be discussed. It was forgotten that with this scoop, the journalists destroyed the castle that built by themselves with lots of efforts. They should understand that its time to quit from the excluding mentality."¹⁴

The reasoning in the independent media that "without the Serbian repression, there would be no peaceful resistance, which disappointed and depreciated the citizens of Kosovo"¹⁵ oriented that media toward objectivity and reality.

The concern for the trend of some media politically influenced to deepen this phenomenon was evidenced in the media's area, where it was claimed that "The Kosovo war has re-dimensioned everything in the Albanian life; the quarrel in politics, the will to eliminate the rival is strong. The news is that the politicians in both sides of the boundary are using the future of Kosovo as weapon, with they would like to kill the political rival by neglecting the fact that this war kills the future of Kosovo. The politics and the Tirana press is the arena where the DLK and he KLA are fighting."¹⁶

The fact that the "Albanian media is divided into advocates of Thaci and Rugova"¹⁷ was not only an indicator of this phenomenon with its negative impact in the public opinion, but also of dependence of media from the politics.

The independent media drew the attention that "the pages of the Albanian press blacken from the attacks and the countercharge. The fate of the nation lays on these. The war unites the politics of one nation."¹⁸

¹² Qose R. 1992 "Shqiptaret kunder vetvetes" Vol Tetor – Dhjetor N 253

¹³ Bukoshi B. 2003 "Kohe sfiduese, kohe e rende per Kosoven" Gazeta Shqiptare Vol Dhjetor N 2569

¹⁴ Shkupi S. 1999 "Arma politike e fyerjes ne shtypin shqiptar"Albania Vol Prill N 143

¹⁵ Palokaj A. 2011 "Kur dhe kush e çliroi Kosovën?" 14 Qershor www. gazetatema.online (retrived June 2011)

¹⁶ Bejtja A. 1999 "Politikane bashkohuni" Gazeta Shqiptare Vol Prill N1235

¹⁷ Xhaferri A 1999 "Erozion" Gazeta Shqiptare Vol Maj N 1260

2. Case of Kosovo. Experience in the theoretical and practical field for the Albanian Media

2.1. Editorial policies in media in Albania (during both periods)

Regarding the problematic of years 1981-1990, Kosovo's agenda setting was decided by the high level party instances, since media was entirely controlled by the party and there was a lack of editorial independence and alternative media. The pluralistic press gave new dimensions to the agenda setting of Kosovo case. For the first time, there exists the independence of media and free thinking. It should be stated that the editorial independence regarding the Kosovo issue was better reported in comparison to other inner issues in Albanian politics.

2.2. Sources of Albanian media during this period.

The information sources in the time of communist system were limited. The sources were official from the Central Committee. The foreign agencies and western press selected to publish or to broadcast were controlled by the Central Committee, too. This phenomenon was related with the isolation of Albania and a closed media, without correspondents and with a full control of the sources that it used.

While in the period of democratic transformation, media operated with considerable sources of information about Kosovo. But the representatives of the Albanian media have not been delegated in Kosovo neither before nor during the war. Therefore, there was a lack of the direct source from the ground. However, the Albanian media used efficiently the indirect and the foreign source but the Albanian reporters were in positioned in the north part of Albania. All the Albanian sources have covered 60% of information on TV channels, meanwhile "during 77 war days 37% of Albania's newspaper pages, 35% "Koha Jonë", 33% "Gazeta Shqiptare" and 30% "Shekulli" newspaper were overwhelmed by Kosovo events, by remarking in this way a kind of record. In this space, the Albanian's sources consisted over 60% of information."¹⁹ The direct continuous information was generally supported in this sources :

The NATO's sources, the Albanian Liberation Army sources too and foreign media that was located in Beograd.

The Albanian Media located on the North part of Albania.

The Serbian sources have been used too. "One of the positive sides of the Albanian coverage Kosovo war was without any doubt the fact that what was happening in Belgrade was not hidden to Albanian public. NATO-s bombs, destroyed bridges, murdered and wounded, Drashkovic's and Miloshevic's interviews on TV were broadcasted without any hesitation. The sources were international media, mainly the BBC and the CNN."²⁰

The Russian and Chinese Media have been referred generally for official information, diplomatic negotiation; Greek Media was rarely referred by the Albanian Media. But the Albanian Media (press and TV broadcasting) didn't include themselves on the international debate to NATO' intervention in Kosovo's war and publishing alternatives opinion of Noam Chomsky, Philip Hammond, Diane Johnston, Daniel Pearl etc were forbid, too. This source missed in Albanian media.

¹⁹ Lani R.L 1999 "Pamje nga Tirana : Lufta dhe Paqja" Media Shqiptare (Korrik – Shtator) Botim AMI, IREX, SOROS Tirane f 63

²⁰ Ibid

3. The role of the Albanian media during this period.

3.1. War media - Challenge for the Albanian media.

Media gains a new experience during this situation. Coverage's war and refuge's crises was unprecedented for the Albanian media. Even before the Albanian media faced emergent situation such as that of natural character, internal social turnings, such that of 90-92, the collapse of pyramid scheme of 97 and the disorder of 98. However, there was not a first hand experience in covering the war events. Therefore, the difficulties that the Albanian media faced were more sensitive in comparison to the foreign ones. In this relatively long but very important period, significance steps were undertaken in two directions: that of professional improvement and that of the consolidation of the Albanian media attributes.

For the first time Albanian media treated a new aspect that of the media in times of war that was unknown before. There is a open way to establish a new relationship between the professional and national interests.

3.2. National or realism of the Albanian media?

The Albanian media is generally presented without any nationalism dimensions. In the relationship realismnationalism, realism prevailed. This happened due to some reasons that are related to the character of the Albanian population and its historical development. In the history of the Albanian population, no records are found to proof the existence of the nationalist ideologies. From the other hand, the 50 years Marxism-Leninism ideology with the principles of internationalism left its traces. This has result that media as part of the social thinking to be released from the nationalism. However, this does not exclude the existence of some nationalist voices, especially in the newspapers of minor political subject for political and public capital.

3.3. Attribute of the Albanian media during Kosovo's War.

Media played an important role in the public opinion in raising awareness towards major events for Albania and in general for the national case. This fact raised the interest of the citizens towards media in general because it contributed to the real evaluation of the situation through the variety of the factual information and experimenting of all types.

Independent of the affiliation, state or party, media expressed the protection of the national interest for the entire range of the thematic from the news, analytical writing and the selection of the foreign press.

At the same time, it showed that it fought for the protection of the democratic human values that are primarily the respect of the human fundamental rights.

The Albanian media used its fourth power to show to the political factor in Kosovo they need unifying in that time to be estimable by its people.

Humanitarian attribute is another dimension of the Albanian media of that time. "In Albania, the media became involved actors through the use of almost all elements of electronic media, public as well as private. They beamed out programs in which family members reunited and got in contact with relatives that were fighters in the Kosovo Liberation Army. Many foreign and especially Western correspondents came to Albania, which created good conditions for presenting the Albanian issue in the international news flow. The news was received directly through the foreign journalists and media teams sent there and imparted and less through the state news agency ATA.²¹

It also has the merits in the creation of the other image of an Albanians and of a society with values that focuses on the human that would affect the future.

²¹ Yamila Milovic "The media and human rights in Kosovo" The International Council on Human Rights Policy. Geneva, 27- 28 November 2000 www.ichrp.org/files/papers/45/106 Kosovo Media Reporting Human Rights Issues Milovic Yamila 2000.pdf (retrieved December 2005)

4. Data and Methodology

The first chapter analyses the Kosovo problematic in the media from the time period from 1981 to 1999. The raised questions for study purposes regard the coverage of the relationship between the ideological criteria of media of that time, state-political pragmatism and the reality in Kosovo, by particularly focusing on the their priority request for the right of self-determination and the way for its fulfillment, applied by the Albanian leadership there. The study method relies on the media analysis in respect to the chronologic development of the events of the respective period. The second chapter describes the Kosovo case as an experience in the theoretical and practical field for the Albanian Media. The information sources of the Albanian media, the identification and their balance, the selection of foreign media's sources and the respect of the ethical and professional standards are in the focus of this chapter. The analysis and the comparison as a study method in this chapter assisted for the clarification of the relationship between sources and propaganda, transparency and manipulation of the foreign source used by the media in Albania. The third chapter describes the role of the media in Albania public opinion through its reporting for that situation during this period, in particular in reporting the war in Kosovo. The comparison and conclusion a study method in this chapter helped to show what kind of attribute defined the Albanian media during that time.

5 Conclusions

1. In general the Albanian media presents itself coherent in its content regarding the Kosovo case. Self determination question remained a crucial issue even in the pluralist media.

2. Freedom of media and pluralist media brought wide information but the analytical level regarding the developments in Kosovo remained limited. The dependence from the foreign analysis characterizes the media of that time.

3. Media exerted its duty in difficult conditions in Albania. The severe political conflicts in Albania and Kosovo was somehow reflected in media both in the state, party and independent media.

4. The war in Kosovo marked for Albanian media its inclusion for the first time in a real fighting ground campaigned by one of the greatest humanitarian crises with high number of refugees. The country's media legislation referring to the war circumstances or humanitarian crises was insufficient in this regard.

5. This period enriched the Albanian media. The gained experience gave to Albanian media the dimension of a journalism school.

References

Books

Balabanova E. 2007 "Media Wars and Politics. Comparing the Incomparable in Western and Eastern Europe" Ashgate Publishing Company USA

Benson L. 2001 "Yugoslavia : A Concise History" Palgrave NY p 146

ASHSH 2009 "Historia e Popullit Shqiptar" Toena Tirana V 4 f 389

Kurspahic Kemal 2003 "Prime time crime – Balkan media and peace" United States Institute of Peace Washinton DC.

Janjic D. 1993 "National Identity, movements and Serbian and Albanians nationalism" D.J "Interethnic Relations in Serbia/Yugoslavia: Alternatives for the Future" CEU Budapest p 122

Kraja M. 1995 "Vite te Humbura" Eurorilindja Tirana f 113;167

Lalaj A. 1994 "E drejta per vetvendosje dhe shqiptaret e Kosoves" Studime Historike N 3-4 f 114

Mertus J. 1999 "Kosovo : "How Myths and truths started a War" University of California Press; Ltd London England

Journal

Lani R. 1999 "Pamje nga Tirana : Lufta dhe Paqja "R.L Media Shqiptare (Korrik – Shtator) AMI, IREX, SOROS Tirana f 63

Newspaper;

Bejtja A. 1999 "Politikane bashkohuni" Gazeta Shqiptare Vol Prill N 1235

Bukoshi B. 2003 "Kohe sfiduese, kohe e rende per Kosoven" Gazeta Shqiptare Vol Dhjetor N 2569

Cani A. 1999 "Lufta bashkon politiken" Gazeta Shqiptare Vol Maj N 1276

Qose R. 1992 "Shqiptaret kunder vetvetes" Rilindja Demokratike Vol Tetor - Dhjetor N 253

Shkupi S. 1999 "Arma politike e fyerjes ne shtypin shqiptar" Albania Vol Prill N.143

Korabi A. 1981 "Kerkesa Kosova Republike nuk mund te shuhet as me dhune, as me teorizime boshe" Zeri i Popullit Vol Korrik Shtator 1981 N 218 (10331)

Xhaferri A. 1999 "Erozion" Gazeta Shqiptare Vol Maj N 1260

Newspaper On Line

Chomsky N. 1999 "Kosovo Peace Accord" Z.Magazine chomsky.info/articles/199907

Chomsky N.2001 "A Review of NATO's War over Kosovo". Magazine www.chomsky.info/articles/200005--htm

Palokaj A. 2011 "Kur dhe kush e çliroi Kosovën?" www. gazetatema.online (retrived June 2011)

Documents On line

Cubrilovic V. "*The expulsion of the Albanians*". Memorandum presented on March 7, 1937 in Belgrade www.trepca.net/english/2006/the_expulsion_of_the_albanians_by_vaso_cubrilovic_memorandum_in_193 7.html

Milovic Y M *"The media and human rights in Kosovo"* The International Council on Human Rights Policy. Geneva, 27- 28 November 2000 www.ichrp.org/files/papers/45/106 Kosovo Media Reporting Human Rights Issues Milovic Yamila 2000.pdf (retrieved December 2005)

Th Annual South-East European Doctoral Student Conference

SEERC is a research centre of the International Faculty of the University of Sheffield, CITY College.

Mission

SEERC's mission is to support the development of South-East Europe by conducting pure and applied research in and for the region. To accomplish this, SEERC employs the existing research capacities of the University of Sheffield and CITY College- International Faculty of The University of Sheffield by facilitating collaborations between their research staff and by developing multi-disciplinary networks of researchers from across South-East Europe.

SEERC contributes to knowledge capacity building:

- By promoting a dialogue among SEE countries related to the development of the SEE region.
- By providing a forum for researchers in the SEE region.
- By disseminating research findings to academics, professionals, NGOs, and policy-makers through conferences, advising, published research papers, monographs, and working papers.
- By developing multi-disciplinary research perspectives through workshops and seminars.
- By promoting the integration of researchers from universities and research centres in the SEE region into networks of excellence.







