

**Proceedings of the 1st Annual SEERC
Doctoral Student Conference**

**Proceedings of the 1st Annual SEERC
Doctoral Student Conference**
Thessaloniki, July 10, 2006

Edited by:
Iraklis Paraskakis
Ioanna Stamatopoulou
Ognen Paunovski

Se auth th selida 8a einai to ISBN etc.

Contents

<i>Preface</i>	<i>vii</i>
<i>Session 1: Business and Economics</i>	<i>1</i>
Implementation of Green Supply Chain Management in South-East Europe – A Literature Review	
Andreas Baresel-Bofinger, Panayiotis H. Ketikidis, S.C. Lenny Koh, John Cullen	3
The Role of Producers and Consumers in Managing Reverse Supply Chains	
Maria Chatzipanagioti, Eleftherios Iakovou	13
Alignment of Core Ideology and Organizational Values with Strategy and Tactics	
Vassilis Kafatos, Leslie Szamosi, John Kawalek	27
Cluster Policy and Entrepreneurship Promotion in Transition Countries in SEE: The Case of Bulgaria, Republic of Macedonia (FYROM) and Serbia	
Aleksandar Karaev, S.C. Lenny Koh, Leslie T. Szamosi	31
Wealth Effects of Mergers and Acquisitions in the Greek Banking Sector	
Aikaterini Markidou, Eftychia Nikolaidou, Simone Mateut	41
Development of Clusters in South-East Europe: the Case of Bosnia and Herzegovina - a Literature Review	
Amira Vejzagic-Ramhorst, Panayiotis H. Ketikidis, S.C. Robert Huggins	53
The Impact of Technological Transfer from Foreign Direct Investment (FDI) – Preliminary Results from Albania	
Mamica Skenderi, Phil Wright, Costas Alexiou	69
<i>Session 2: Information Technology and IT Policy</i>	<i>95</i>
Pedagogical Developments for Teaching and Learning using VLEs: A Phenomenographic Inquiry into Greek University Teachers’ Conceptions and Approaches	
Petros Lameris, Phil Levy, Iraklis Paraskakis	97
Querying and Visualizing the Semantic Web	
Georgios Meditskos, Efstratios Kontopoulos, Nick Bassiliades	109
A Model Framework to characterise the NEHRS Preparedness: Validation Phase	
Leonidas Orfanidis, Barry Eaglestone, Panagiotis Bamidis	120
Investigation of Emergent Phenomena within Complex Adaptive Systems	
Ognen Paunovski, George Eleftherakis, Tony Cowling	136
A Hybrid Formal Approach for Modelling Change in Multi-Agent Systems	
Ioanna Stamatopoulou, Petros Kefalas, Marian Gheorghe	146
Modern Applications of Machine Learning	
George Tzanis, Ioannis Katakis, Ioannis Partalas, Ioannis Vlahavas	156
An Analytical Strategic Model for the Configuration of a Closed-Loop Supply Chain Network	
Anastasios Xanthopoulos, Eleftherios Iakovou	166

<i>Session 3: Political and Social Issues</i>	175
A Theoretical Analysis of the Europeanisation Impact in Macedonia	
Ismail Aydil, Andrew J. Taylor, Miltiadis Sarigiannidis.....	177
Effects of Europeanisation on Social Policies of Slovenia and Macedonia: Convergence vs. Disparity?	
Maja Gerovska Mitev	189
Gaining or Losing Identities: An Investigation into Perceptions of the Changing Status of English as a Lingua Franca in Bulgaria and Serbia	
Sarah Hannam.....	203
Integrated Border Management: Constructing Cross-Border Spaces	
Despina Syrri	210
Talking about Migration in Central Northern Greece: The Employment of the Construct of Greek Emigrants	
Maria Xenitidou, Paul White, Sean Homer.....	224
EU Police Missions: from Conflict Management to Institution Building in the Western Balkans	
Islam Yusufi, Andrew Taylor, Miltos Sarigiannidis	236
<i>Session 4: Psychology and Culture</i>	243
Visuospatial Attention in Dyslexia	
Margarita Diakou, Ana B. Vivas, Rod I. Nicolson.....	245
Autism, Communication Impairment, and Gastrointestinal Symptoms	
Sophia Dimopoulou, Efrosini Kalyva, Michael Siegal	256
Consuming Nationalism: Contemporary Nationalism in Turkish Popular Culture	
M. Sinan Gökçen, Sean Homer, Caroline J. Oates.....	261
In Search of the Invisible Balkan Cinema	
Efthimios Hatzis, Erica Sheen, Sean Homer	274
Modulation of IOR as a Function of SOA Value	
Fenia Kalogeropoulou, Ana Vivas, Peter Woodruff	285
A Preliminary Study of Greek Adolescents' Estimates of Smoking Prevalence: Beyond the Effects of False Consensus	
Lambros Lazuras, Angelos Rodafinos, J. Richard Eiser	293

Preface

The 1st Annual SEERC Doctoral Student Conference (DSC06) took place on July 10 in Thessaloniki, Greece, and was organised by the South-East European Research Centre.

The aim of the conference was to initiate an exchange of knowledge between young researchers and to help establish a network of scholars currently undertaking research in South-East Europe. 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.

Twenty-five young researchers mainly from SEERC Doctoral Program as well as other institutions in South East Europe (SEE) presented their work in the conference, which was well attended with approximately fifty registered participants.

The event commenced with welcome speeches from Dr. Iraklis Paraskakis (Director of SEERC Doctoral Program) and Mr. Nikos Zaharis (SEERC Director) who talked about the research challenges faced by South-East Europe. These were followed by a speech from Dr. Nikos Dimitriadis who shared his recent experience on the path to completing his PhD.

The scope of the conference was multi-disciplinary spanning throughout the areas in which SEERC is doing active research and therefore it was divided into four parallel sessions:

- Business and Economics,
- Information Technology and IT Policy,
- Political and Social Issues,
- Psychology and Culture.

The Business and Economics as well as the Information Technology and IT Policy sessions included seven paper presentations each. Five papers have been presented in the Political and Social Issues session while the participants also had the opportunity to attend a workshop entitled “Nationalism, Ideology and Balkan Cinema”. Finally, the session on Psychology and Culture featured 6 paper presentations as well as an open discussion at the end.

One of the objectives of the conference has been to provide an opportunity for PhD students to receive advice from experts in their chosen field of research. This would not have been accomplished without the participation of the invited discussants. We would like to thank Prof. Siniša Zarić (Director, Centre of International Studies, Faculty of Economics, University of Belgrade), Assoc. Prof. Yannis Hajidimitriou (Department of Business Administration, University of Macedonia), Prof. Ioannis Vlahavas (Department of Informatics, Aristotle University of Thessaloniki), Asst. Prof. Nick Bassiliades (Department of Informatics, Aristotle University of Thessaloniki), Assoc. Prof. Asteris Huliaras (Department of Geography, Harokopion University of Athens), Dr. Antonis Sapountzis (Department of Primary Education, Aristotle University of Thessaloniki), Dr. Vassilis Barkoukis (Department of Physical Education and Sport Science, Aristotle University of Thessaloniki) and Dr. Efrosini Kalyva (Department of Psychology, CITY College) for accepting our invitation and providing their valuable feedback.

Closing we would like to also thank the organising committee, the authors of the papers, all the presenters and participants and our colleagues at SEERC that contributed in making DSC06 a successful event. We are looking forward to the announcement of the 2nd conference.

Iraklis Paraskakis
Ioanna Stamatopoulou
Ognen Paunovski

SESSION 1:

BUSINESS AND ECONOMICS

Implementation of Green Supply Chain Management in South-East Europe – A Literature Review

Andreas Baresel-Bofinger¹, Panayiotis H. Ketikidis¹, S.C. Lenny Koh², John Cullen²

¹SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, {abofinger, pktikidis}@seerc.info

²The University of Sheffield, Management School, 9 Mappin Str, Sheffield S1 4DT, UK, {S.C.L.Koh, john.cullen}@sheffield.ac.uk

Purpose – With the recent economic progress and the increasing attractiveness of the region of South-East Europe (SEE) and the growing shift of production facilities from Western European countries to the regions of Eastern and South-East Europe the concept of Environmental (Green) Supply Chain Management (GSCM) will play a growing role in these countries. This research aims in its initial stage to give an overview of the existing literature about the various aspects related to green supply chain management with the intention to examine their application for the region of SEE. **Design/methodology/approach** – The above purpose will be achieved through an extensive literature review surveying professional publications and scholarly research in the areas of green supply chain management, sustainability and waste management with reference to EU efforts and initiatives and the current situation in SEE. **Findings** – Summarising relevant literature this paper will provide a base for future discussions about the application of green supply chain management in the economies of SEE. The findings identify major problems to be addressed in this process. **Research limitations/implications** – As green supply chain management is still a rather new subject in the region of SEE, there is limited research about its appropriate application. An implication of this paper is the need for more studies in this field. **Practical implications** – This work can serve as a source of valuable information for researchers and other parties interested in the development of mechanisms of green supply chain management in SEE. **Originality/value** – This work extends the research of green supply chain management into the area of South-East Europe.

Keywords

Green supply chain management, sustainability, waste management.

1. Introduction

1.1 Awareness of environmental performance

The interest in an organisation's environmental performance has been steadily growing over the last decade [1-6]. Research has explored many aspects of firms' approaches towards the natural environment and looked at possible responses to the challenges imposed by tighter environmental regulations, restricted natural resources, customer demands, competitive pressures and ethical responsibility [7-16]. Companies all over the world are globalising more and more of their operations, resulting in increasing value chain complexity entailing physical, financial and information flows. Western European manufacturers have a strong focus on cost reduction and outsourcing of operations. They are moving production to low labour cost regions such as South East Asia and China but also to the region of Central, Eastern and South-East Europe. The increased activity in manufacturing and goods transportation means also increased generation of waste in those regions, produced by industry

and households. This creates also a threat for the environment. Manufacturing and production enterprises will have to develop ways in which industrial development and environmental protection can symbiotically coexist [17]. The challenge is to redefine the basic structure of the entire supply chain. The globalisation of supply chains extends requirements for environmental-friendly operations and strategies also to “newcomers” from the transient and emerging economies.

1.2 The region of SEE

While the regions of South East Asia and China have been put in the focus of latest research concerning greening the supply chain [18-22] also the countries neighbouring the European Union, such as the region of South East Europe (SEE), will have to reconsider their standpoint regarding corporate environmental strategies if they want to take advantage of their proximity to the European Union and be part of a supply network that is characterised by a greater awareness of environmental protection. The impact of the growing EU environmental legislation will be felt across almost every sector of the European economy and all its vicinity. Businesses responding too slowly or ineffectually are likely to incur significant cost increases, while businesses responding in a timely and structured manner can expect to enjoy not only a competitive advantage, but also a potential reduction in costs [23]. The countries of South-East Europe, who are already part of the European Union such as Greece and since 2004 also Cyprus and Slovenia, and the candidate countries Bulgaria and Romania, as well as Croatia and Turkey, and other countries of the region, such as Albania, FYROM, Serbia and Montenegro, and Bosnia-Herzegovina have an interest and incentive to comply with the environmental guidelines being implemented in Europe if they wish to become competitive in an extended European market.

1.3 Aim and objectives of the research

In a first stage this research attempts to review the existing literature in order to identify the major topics related to green supply chain management and to explore the challenges to be faced in the attempt to facilitate implementation of GSCM in the region of SEE.

2. Research methodology

Performing an extensive state-of-the-art literature review a taxonomy will be created identifying the contributions and gaps of the various existing approaches. The base for this taxonomy is a taxonomy table containing a number of categories as shown in example Table 1:

A continuous review of this taxonomy table will gradually allow drafting a web of interconnected fields of research related to the area of GSCM.

3. Definition

The definition of green supply chain management evolves from the understanding of supply chain management.

3.1. Supply Chain Management

There are various attempts in literature to define supply chain management (SCM). [24-26] As a typical example may serve the one given by Handfield and Nichols [25] who describe SCM as all those activities associated with the flow and transformation of materials from the raw extraction of materials through to the consumption of goods and services by an end user, along with the attendant information flows, both up and down the supply chain.

Author(s)	Bowen F E, Cousins P D, Lamming R C and Faruk A C	Zhu Q, Sarkis J and Geng Y
Title	The role of supply management capabilities in green supply.	Green supply chain management in China: pressures, practices and performance.
Publication Type	Journal article	Journal article
Publication name	Production and Operations Management	International Journal of Operations & Production Management
Volume (No)	10(2)	25(5)
Pages	174-89	449-68
Date/Year	2001	2005
Thematic Topic	Analysis of the role of supply management capabilities in green supply	Evaluating GSCM drivers, practices and performance in China
Concept / Theory	Focusing on development of firm's specialised internal resources; relationship with supplier	Defining status-quo of GSCM in Chinese manufacturing enterprises
Methodology	Two-phase survey of 70 operating units in public companies, broad cross-section of industries, "clean" and "dirty", semi-structured interviews, questionnaires, LISREL analysis	Survey questionnaire with 54 items, 314 responses; exploratory factor analysis to make groupings; comparative analysis with previous research
Issue / Challenge	How to develop appropriate internal supply management capabilities in order to effectively implement green supply initiatives	Reaction of a transient economy to environmental pressures
Country	United Kingdom	China
Contribution	Development of a basic predictive model; Supply management capabilities are jointly developed by proactive corporate environmental attitude and more strategic purchasing and supply management approach; SMC facilitate implementation of product-based green supply but not greening the supply process <> contrary to current research	Definition of factors for GSCM analysis; higher awareness in Chinese enterprises through external pressure but no translation into strong GSCM implementation
Supporting	Drumwright 1994; Cramer 1996; Green, Morton and New 1996; Lamming and Hampson 1996; Min and Galle 1997; Noci 1997	Russel 1998
Contradicting	partially Carter 1998	Min and Galle, 1997: US firms: potential liability and disposal cost-Chinese firms: regulatory issues; Zsdisin and Hendrick 1998: Western firms: investment recovery, Chinese firms: less important
Impact	Supporting evaluation and development of corporate readiness for green behaviour	extending GSCM research into China
Shortfall	small sample; unusual sampling strategy; focusing only on purchasing function	little previous research about China, little comparison; limited sample from special economic zones
Future direction	Role of firms internal capabilities in supporting environmental management in other process and functional areas besides purchasing and supply; different capabilities needed for different green initiatives?; > Resource-based perspective	Relationship between identified factors

Table 1 Taxonomy table example for Green Supply Chain Management

3.2. Green Supply Chain Management

A helpful definition of green supply chain management (GSCM) is given by Hervani et al [27] who characterise GSCM as a composition of green purchasing, green manufacturing /green materials management, green manufacturing, green distribution/marketing and reverse logistics.

4. Major topics of research

At the current stage this literature review focuses on the past approximately ten years. Research in green supply chain management and related areas has covered a wide variety of topics. Also some major case studies both related to specific industrial sectors and geographic areas have been conducted. Figure 1 gives a preliminary overview of the various topics of research grouped in related thematic fields.

Among the various identified research fields there is a group of thematic topics of special interest for researchers dealing with the firms' operational level focusing on the relation between firms' proactive behaviour [28-32] and innovation level [33-38], environmental cost accounting [28,30,39-47] and performance measurement [2,29,31,37,43,48-51].

4.1 Proactivity

Sadgrove [52] classifies firms' environmental behaviour along a range from reactive to proactive approaches. A proactive approach entails the ability to foresee future trends in society and regulations and adapt operations, processes, and products accordingly in order to avoid negative impacts on the environment. [47] Russo & Fouts [46] find a positive relationship between proactive environmental strategies and organisational performance results when firms develop complex capabilities. According to Aragón-Correa [9] strategically proactive firms approach natural environment both correctively and proactively. Shrivastava [53] points out the difficulty for less proactive firms to invest in new technologies.

4.2 Innovation

Starik and Rands [38] stress the need for sustainable firms to design their internal processes to enable employees to get involved in innovative measures that are directed towards sustainability. Bringer and Benforado [54] add the point that it is important for environmentally advanced firms to invest in R&D and to support natural environmental aims throughout the entire organisation. Angel de Brio and Junquera [55] hint to the obstacles faced by SMEs when trying to adopt innovative measures. Geffen and Rothenberg [48] examine the role of suppliers as an important source for innovation, while Porter and Van der Linde [56] find competitive and regulatory pressures as the main motivators for innovation.

4.3 Environmental cost accounting

Nehrt [57] considers environmental consciousness as a source of competitive advantage in international markets. Similarly, Hansmann and Claudia [58] see implementation of GSCM as a new opportunity for competition and a new way to add value to core business programmes. Also Hutchison [59] assigns to GSCM the major role for addressing key factors for competitive advantage through environmental performance. Van Hook and Erasmus [60] see profit and market share objectives met through lowering environmental risks and raising ecological efficiency.

Klassen and McLaughlin [29] make a clear statement in their research that proactive corporate environmental strategies result in improved financial performance and lead to environmental and competitive improvements in association with the development of certain strategic managerial and manufacturing processes. Sharma & Vredenburg [47] explain in the framework of the resource-based

environmental performance. Alvarez et al [62] discover a positive effect of greening the supply chain on a firm's economic performance. Contrary to that, Bowen et al [40] declare that they could not find short-term profitability and better sales performance as a result of GSCM. In a study on Chinese enterprises Zhu et al [19] confirm that there is no improved economic performance through implementation of GSCM. Hanna et al [63] find in their research a strong relationship between meeting operational goals and staff involvement on environmental management. Sroufe [64] creates a framework with performance indicators and supplier assessment metrics for gaining competitive advantage and reducing risk. Hervani et al [27] give an overview of performance measurement literature and draft an integrative framework for study, design and evaluation for GSCM tools. Kainuma & Tawara [65] construct a multi-attribute utility function of the supply chain and refer to the impact of information sharing. Simpson et al [66] ascertain that customer performance requirements on suppliers have an impact on suppliers' environmental performance. Rao [41] conducts a study on performance measurement in South East Asia commenting on the progress and difficulties of implementing GSCM in that particular region. Harvey and Schaefer [67] discover that external reporting serves as pressure for better performance results.

5. Review of controversial issues

One of the topics finding controversial response through research results refers to the issue if value can be created through an environmentally responsible supply chain. Although there is little doubt that more stringent environmental standards have to be met and many organisations have to devote increasing resources to develop and implement corresponding measures, there is no clear answer to the question if a better environmental performance results also in a better economic performance.

In this context falls also the discussion if an environmental proactive approach is a guarantee for better economic performance. It is arguable if a company going beyond simple compliance can achieve a sustainable competitive advantage.

A related topic concerns performance measurement. It is still to be tested which is best way to measure environmental performance. It is debated which metrics and which tools are best suitable for this task.

Thus the development of a conceptual model linking green supply chain management, competitiveness and economic performance has not found a final answer yet. [66,68]

Researchers are still arguing what constitutes an environmentally excellent supply chain. There is a continuous challenge to find out what are the crucial factors and best implementations to achieve environmental excellence in a supply network.

6. Challenges for environmental research topics towards South-East Europe

While environmental decline has taken place in the region of South-East Europe during the time of Soviet authority, economic hardship did not allow the development of new techniques in the field of protection of the natural environment after the end of the cold war. Still many of the newly independent states have big financial difficulties to undertake the efforts necessary to make industrial processes more environment-friendly in spite of a recent shift of attitude. The enlargement of the European Union and the interest in stability in the neighbouring states has put the development of the economy in the region of SEE into focus. The development of green supply chain management in countries that are on the way to play a stronger economic role in the global market without having yet adopted a clear strategic approach toward the natural environment will gain increasing importance in the near future. SEE, which is in close proximity to the European Union, is one of the regions that faces the challenge to comply to rigid regulations, customer demands and competitive pressure if it intends to take a more active part in the European and international supply chains.

Issues that will have to be addressed are the question how the gap between awareness of the necessity to introduce green supply chain management and its practical implementation can be successfully bridged and how support from management and government has to be shaped. It may be worth examining if the encouragement of voluntary initiatives, as it has been done in Asia, could be a promising path to go in the political, economic and cultural context of SEE. As in other economies the examination of the increasingly important topic of performance measurement combining environmental, operational, and economic level in order to show a clear linkage between good environmental performance and improved economic performance may prove to be a useful way to persuade organisations in the region of SEE to adopt GSCM. The development and improvement of good performance measurement tools for green supply chain management is a promising field of further research.

Based on a number of topics suggested for further research by Hervani et al [27] in the field of performance measurement tools of green supply chain management the following challenges can be identified:

Research must examine the business and environmental results of a GSCM performance measurement system and their impact within the organisation, industry, and society in SEE. In this framework further investigation should go into inter-organisational agreement on performance management and measurement. Different types of metrics need to be developed, potential designs and tools of a GSCM/PMS. In order to determine which type of performance measurement systems work best in which sector, industry-specific research is recommendable. The development of appropriate data and information with respect to GSCM is an important topic for further studies. The question is to be answered which roles new technologies including information technology can assume in this process. It should also be scrutinized how appropriate the current tools and management of supply chain management are for incorporating environmental management dimensions. From a global perspective the question is how the firms in the SEE can be incorporated efficiently into global supply chains in respect to environmental issues. To what extent can other management tools, such as life cycle analysis, in combination with performance measurement help implement GSCM.

The common understanding is that GSCM creates and improves synergy and efficiency among business partners and helps to enhance their environmental performance, reduces waste and supports cost savings measures. In spite of these apparent advantages companies in SEE will have to be shown a clear link between the adoption of such measures and resulting better economic performance and competitiveness. Research should test an empirical link between such efforts and subsequent improvements in competitiveness and economic performance. In this context research should also investigate the question what is environmental excellence in a firm's supply chain management response to all the environmental pressures it faces. The definition of environmental operating models, operational objectives, and new supply chain processes is needed for the firms of the region of SEE. Research may establish a relationship between the quality of a supply chain response and the extent of competitive advantage. It can provide to firms a framework that addresses environmental pressures, and show them a way how to assume a proactive approach to create supply chains resulting in improved profitability and environmental sustainability.

7. Conclusion

This work has given a preliminary review of the existing literature over the last decade concerning green supply chain management and has identified some of the major current research topics in that area. The paper has made the attempt to point out controversial issues and the need for further research in these thematic fields. With society's growing understanding that the environment should not have to pay the price for economic growth research will continue to examine the relationship between business and environment on a multitude of layers. Transition economies such as the region of SEE necessarily will have to focus on environmental issues if they want to compete in the global supply chain networks. Being a latecomer in the field firms from that region have the advantage to be able to avoid mistakes made by their competitors in the past and apply modern practices and tools of corporate environmental management on strategic and operational level.

This literature review needs to be extended to cover a larger period of time to serve as a base for a further in-depth research focusing on the particular conditions of selected countries of the region of SEE and evaluating various models and mechanisms to implement GSCM. The result of such a research should help firms of the region to gain insight how to develop an environmental approach that allows them to become more competitive and successfully integrate themselves into global supply chains.

References

- 1 Shultz C J II, Holbrook M B. Marketing and tragedy of the commons: a synthesis, commentary and analysis for action. *Journal of Public Policy and Marketing* 1999; 18(2): 218-29.
- 2 Sarkis J. Manufacturing strategy and environmental consciousness. *Technovation* 1995; 15(2): 79-97.
- 3 Hick S. Morals make the money. *Austrian CPA* 2000;70: 72-3.
- 4 Angell L C, Klassen R D. Integrating environmental issues into the mainstream: an agenda for research in operations management. *Journal of Operations Management* 1999; 17(5): 575-98.
- 5 Geyer R, Jackson T. Supply loops and their constraints: the industrial ecology of recycling and reuse. *California Management Review* 2004; 46(2): 55-73.
- 6 Gupta M. Environmental management and its impact on the operations function. *International Journal of Operations & Production Management* 1995; 15(8): 34-51.
- 7 Faruk A C, Lamming R C, Cousins P D and Bowen F E. Analyzing, mapping and managing environmental impacts along the supply chain. *Journal of Industrial Ecology* 2002; 5(2): 13-36.
- 8 Banerjee S. Corporate environmentalism - The construct and its measurement. *Journal of Business Research* 2002; 55(3): 177-191.
- 9 Aragon-Correa J A. Strategic proactivity and firm approach to the natural environment. *Academy of Management Journal* 1998; 41(5): 556-67.
- 10 Bowen F E, Cousins P D Lamming R C and Faruk A C. Horses for courses: explaining the gap between theory and practice of green supply. *Greener Management International* 2001; 35:41-60.
- 11 Buysse K, Verbeke A. Proactive environmental strategies: A stakeholder management perspective. *Strategic Management Journal* 2003;24(5): 453-470.
- 12 Hanna M D, Newman W R. Operations and the environment: an expanded focus for TQM. *International Journal of Quality & Reliability Management* 1996; 12(6): 38-53.
- 13 Sarkis J. Manufacturing's role in corporate environmental sustainability: concerns for the new millennium. *International Journal of Operations & Production Management* 2001; 21(5/6): 666-85.
- 14 Melnyk S A, Sroufe R P and Calatone R. Assessing the impact of environmental management systems on corporate and environmental performance. *Journal of Operations Management* 2002; 21(2): 329-51.
- 15 Carter C R. Ethical issues in international buyer-supplier relationships: a dyadic examination. *Journal of Operations Management* 2000; 18(2): 191-208.
- 16 Zaklad A, McKnight R, Kosansky A and Piermarini J. The social side of the supply chain: align three factors, and hitting the jackpot is a sure bet. *Industrial Engineer* 2004; 36(2): 40-4.
- 17 Beamon B M. Designing the green supply chain. *Logistics Information Management* 1999; 12(4): 332-42.
- 18 Anbumozhi V, Kanda Y. Greening the production and supply chains in Asia: Is there a role for voluntary initiatives? IGES Kansai Research Centre KRC 2005; 6E
- 19 Zhu Q, Sarkis J and Geng Y. Green supply chain management in China: pressures, practices and performance. *International Journal of Operations & Production Management* 2005; 25(5): 449-68
- 20 Christmann P, Taylor G. Globalization and the environment: determinants of firm self-regulation in China. *Journal of International Business Studies* 2001; 32(3): 439-58.
- 21 Chan R Y K, Lau L B Y. Explaining green purchasing behavior: a cross-cultural study on American and Chinese consumers. *Journal of International Consumer Marketing* 2001; 14 (2/3): 9-41.

- 22 Lo C W, Leung S W. Environmental agency and public opinion in Guangzhou: the limits of a popular approach to environmental governance. *The China Quarterly* 2000; 163: 677-704.
- 23 Monahan S, Van den Bossche P and Harthan C. Shifting your supply chain into reverse. *Kearney Executive Agenda – Ideas and Insights for Business Leaders* 2004 3rd quarter; 7(3).
- 24 Bowersox D J, Closs D J. *Logistical management: The integrated supply chain process*. New York: McGraw-Hill; 1996.
- 25 Handfield R, Nichols, E. *Introduction to supply chain management*. New Jersey: Prentice Hall; 1999.
- 26 Christopher M in Irwin R D. *Logistics and supply chain management*. New York: Financial Times; 2005.
- 27 Hervani A A, Helms M M and Sarkis J. Performance measurement for green supply chain management *BIJ* 2005; 12(4): 330-53.
- 28 Judge W Q, Douglas T J. Performance implications of incorporating natural environmental issues into the strategic planning process: An empirical assessment. *Journal of Management Studies* 1998; 35: 241–262.
- 29 Klassen R D, McLaughlin C P. The impact of environmental management of firm performance. *Management Science* 1996; 42(8): 1199-215.
- 30 Klassen R D, Whybark D C. The impact of environmental technologies on manufacturing performance. *Academy of Management Journal* 1999; 42: 599–615.
- 31 Hart S L, Ahuja G. Does it pay to be green: an empirical examination of the relationship between emission reduction and firm performance. *Business Strategy and the Environment* 1996; 5(1): 30-7.
- 32 Hunt C B, Auster E R. Proactive environmental management: Avoiding the toxic trap. *Sloan Management Review* 1990; 31(2): 7–18.
- 33 Sharma S. Managerial interpretations and organizational context as predictors of firm choice of environmental strategies. *Academy of Management Journal* 2000; 43: 681-97.
- 34 Bansal P, Roth K. Why companies go green: a model of ecological responsiveness. *Academy of Management Journal* 2000; 43(4): 713-48.
- 35 Henriques I, Sadosky P. The determinants of an environmentally responsive firm: an empirical approach. *Journal of Environmental Economics and Management* 1996; 30: 381-95.
- 36 Henriques I, Sadosky P. The relationship between environmental commitment and managerial perceptions of stakeholder importance. *Academy of Management Journal* 1999; 42(1): 87-99.
- 37 Hall J. Environmental supply chain dynamics. *Journal of Cleaner Production* 2000; 8(6)
- 38 Starik M, Rands G P. Weaving an integrate web: Multilevel and multisystem perspectives of ecologically sustainable organizations. *Academy of Management Review* 1995; 20: 908-935.
- 39 Dean T J, Brown R L. Pollution regulation as a barrier to new firm entry: Initial evidence and implications for future research. *Academy of Management Journal* 1995; 38: 288–303.
- 40 Bowen F E, Cousins P D, Lamming R C and Faruk A C. The role of supply management capabilities in green supply. *Production and Operations Management* 2001; 10(2): 174-89.
- 41 Rao P. Greening the supply chain: a new initiative in South East Asia. *International Journal of Operations & Production Management* 2002; 21(6): 632-55.
- 42 Hart S. A natural-resource based view of the firm. *Academy of Management Review* 1995; 20(4).
- 43 Florida R. Lean and green: the move to environmentally conscious manufacturing. *California management review* 1996; 39(1,): 80-105.
- 44 Porter M E. America's green strategy. *Scientific America* 1991; 264(4): 168.
- 45 Marcus A A, Nichols M L. On the edge: Heeding the warnings of unusual events. *Organization Science* 1999; 10: 482–499.
- 46 Russo M, Fouts P. A resource-based perspective on corporate environmental performance and profitability. *Academy of Management Journal* 1997; 40(3): 534-559.
- 47 Sharma S, Vredenburg H. Proactive corporate environmental strategy and the development of competitively valuable organizational capabilities. *Strategic Management Journal* 1998; 19: 729–753.
- 48 Geffen C, Rothenberg S. Suppliers and environmental innovation – The automotive paint process. *International Journal of Operations & Production Management* 2000; 20(2): 166-186.

- 49 Green K, Morton B and New S. Purchasing and environmental management: interaction, policies and opportunities. *Business Strategy & the Environment* 1996; 5: 188-97.
- 50 Handfield R, Walton S V, Sroufe R and Melnyk S A. Applying environmental criteria to supplier assessment: a study in the application of the analytical hierarchy process. *European Journal of Operational Research* 2002; 141: 70-87.
- 51 Frosch R. Industrial ecology: minimizing the impact of industrial waste. *Physics Today* 1994; 47(11): 63-8.
- 52 Sadgrove K. *The green manager's handbook*. Aldershot: Gower; 1993.
- 53 Shrivastava P. Environmental technologies and competitive advantage. *Strategic Management Journal* 1995; 16: 183–200.
- 54 Bringer R P, Benforado D M. Pollution prevention and total quality environmental management. In Kolluru R V. *Environmental strategies handbook: A guide to effective policies and practices*. New York: McGraw-Hill; 1994.
- 55 Angel del Brio J, Junquera B. A review of the literature on environmental innovation management in SMEs: implications for public policies. *Technovation* 2003; 23(12): 939-48.
- 56 Porter M, Van der Linde C. Toward a new conception of the environment-competitiveness relationship. *Journal of Economic Perspectives* 1995; 9(1): 97-118.
- 57 Nehrt C. Maintainability of first mover advantages when environmental regulations differ between countries. *Academy of Management Review* 1998; 23: 77–97.
- 58 Hansmann K W, Claudia K. *Environmental management policies in Sarkis J. Green manufacturing and operations: from design to delivery and back*. Sheffield: Greenleaf Publishing; 2001.
- 59 Hutchison J. Integrating environmental criteria into purchasing decision: value added? in Russel T. *Green purchasing: opportunities and innovations*, Sheffield: Greenleaf Publishing; 1998.
- 60 van Hock R I, Erasmus. From reversed logistics to green supply chains. *Logistics Solutions*. 2000; 2: 28-33.
- 61 Wagner M, Schaltegger S and Wehrmeyer W. The relationship between the environmental and economic performance of firms: what does theory propose and what does empirical evidence tell us? *Greener Management International* 2001; 34: 95-108.
- 62 Alvarez Gil M J, Jimenez J B and Lorente J C. An analysis of environmental management, organizational context and performance of Spanish hotels. *Omega* 2000; 29(6):457-71.
- 63 Hanna M D, Newman W R and Johnson P. Linking operational and environmental improvement through employee involvement. *International Journal of Operations & Production Management* 2000; 20(2): 148-65.
- 64 Sroufe R S. Effects of environmental management systems on environmental management practices and operations. *Production and Operations Management* 2003; 12(3): 416–431.
- 65 Kainuma Y, Tawara N. A multiple attribute utility theory approach to lean and green supply chain management. *Int. J. Production Economics* 2006; 101: 99–108.
- 66 Simpson D, Power D and Samson D. Greening the automotive supply chain: a relationship perspective. *Australian Centre for International Business* 2005 Oct; Working Paper 8
- 67 Harvey B, Schaefer A. Managing relationships with environmental stakeholders: a study of UK water and electricity utilities. *Journal of Business Ethics* 2001; 30(3): 243-52.

Appendix

List of Acronyms

ERP	Enterprise Resource Planning
GSCM	Green Supply Chain Management
PMS	Performance Measurement Systems
RBV	Resource-Based View
SEE	South-East Europe

The Role of Producers and Consumers in Managing Reverse Supply Chains

Maria Chatzipanagioti, Eleftherios Iakovou

Aristotle University of Thessaloniki, Mechanical Engineering Department, Industrial Management Division, 54124 Thessaloniki, Greece, {mchatzip, eiakovou}@auth.gr

The past two decades environmental accountability has become an important issue, with the public, government, and business community showing growing interest in environmental concerns. Environmental laws and regulations have been enacted to hold businesses accountable for their environmental responsibilities and also corporations have taken voluntary initiatives to demonstrate their commitment to environmental consciousness and obligations. Especially for the case of electrical and electronic products, management of waste of such equipment has become a major issue, mainly due to materials value that can be recovered at the end of the products useful life and also to the venturousness of toxic materials used. This theoretical paper aims firstly to identify the environmental laws and regulations regarding industrial electrical and electronic products, secondly to summarize the key points regarding the environmentally related ISO 14001 certification and the implementation of either a take-back or an environmental management system, thirdly to capture the role of manufacturers in the reverse supply chain along with methods to motivate customers participate in product take-back systems, and finally to provide guidelines for future research.

Keywords

Consumer motivation, Environmental management, ISO 14001, Legislation, Take-back systems.

1. Introduction

The past twenty years have been characterized as the era of environmental consciousness. With the growing interest in environmental concerns by the public, government, and business community, environmental accountability has become an important issue. Environmental laws and regulations have been enacted to hold businesses accountable for their environmental responsibilities and also corporations have taken voluntary initiatives, which are an integral part of corporate social responsibilities and demonstrate corporations' commitment to environmental consciousness and obligations [1].

This environmental consciousness brought in a growing interest in the flows of the reverse supply chain, which are generated by the three following sources: a) the consumer, who returns the goods, b) the industry interested in recycling, and c) the government, which promotes practices of this type [2]. According to the above consideration, the consumer returns the goods back to the producer at the end of their useful life, while producers adopt take-back systems in order to comply with laws and regulations, and also take up environmental management systems, because 1) a voluntary initiative, like the ISO 14001 certification, is part of their corporate environmental policy, and 2) plethora of benefits and incentives can be obtained from the implementation of such systems.

In this theoretical paper we firstly identify the environmental laws and regulations regarding electrical and electronic products, secondly we summarize the key points with reference to the environmentally related ISO 14001 certification and the implementation of take-back and environmental management systems, and thirdly we capture the role of manufacturers in the reverse supply chain along with methods to motivate customers participate in product take-back systems.

The rest of the paper is organized as follows. In the following section, we present the role of producers in the reverse supply chain through the existing literature regarding legislation and regulations (subsection 2.1), take-back systems (subsection 2.2), the ISO 14000 model and the ISO 14001 certification (subsection 2.3) and environmental management systems (subsection 2.4). Then, in Section 3, through the existing literature, we classify consumers according to specific characteristics, and we suggest ways to motivate them take back the products to the manufacturer. Finally, we wrap up with summary, conclusions and future research directions.

2. The Role of Producers

As the responsibility for the disposal of products at the end of their useful life shifts through regulation back to the producer, companies have been forced to incorporate these take-back requirements into their strategic planning processes. Therefore, the manufacturer's critical role in the reverse supply chain is not recognized in his decision on implementing a take-back system, since this is obligatory, but exactly in the recognition of legislation and the formulation of the appropriate reverse supply chain, deciding whether he will manage the disposed products or outsource them to a third party (an immiscibly recycling corporation). Furthermore, the producer has to identify his customer's requests (for example the ISO 14001 certification) and finally the benefits that can be obtained from the implementation of an environmental management system. In other words, the producer has to choose between compliance with the minimum requirements (implementation of a take-back system according to the legislation) and formulation and implementation of an integrated environmental management system (proactive green behaviour).

2.1 Legislation Regarding Products' Disposal

Disposed electronic products can cause health concerns because they contain a wide spectrum of hazardous materials, ranging from lead, phosphorus, cadmium, chromium and barium to mercury. The list of human health risks resulting from improper contact with these materials includes breathing difficulties, coughing, choking, respiratory irritation, pneumonitis, tremors, neuropsychiatric problems, convulsions, comas and even death [3]. Moreover, electrical and electronic products are also affected by their rapid technological obsolescence resulting in a significant increase on the waste capacity. Consequently, the legislated prevention of waste generation along with the reuse, recycling and recovery of products at the end of their useful life seems to be necessary in order to reduce the disposal of waste.

Therefore, the European Union (EU) has recently approved a directive on waste electrical and electronic equipment (2002/96/EC) applicable to all EU member states, requiring the recovery of a significant proportion of electronic products at the end of their useful life. These EU regulations require that a certain per capita quantity of electronic products will be recovered by specified deadlines. For hazardous components, recovery targets have been established as high as 80% by weight, while recovery targets for different electronic products range from 50 to 80% by weight [4].

Moreover, the European Union (EU) has also approved a directive on the restriction of the use of certain hazardous substances in electrical and electronic equipment (2002/95/EC). The directive's purpose is to restrict the use of hazardous substances in electrical and electronic equipment and to contribute to the protection of human health and the environmentally sound recovery and disposal of waste electrical and electronic equipment. Specifically, new electrical and electronic equipment put on the market from 1st July 2006 is not allowed to contain lead, mercury, cadmium, hexavalent chromium, polybrominated biphenyls (PBB) or polybrominated diphenyl ethers (PBDE) [5].

The European Union aims through these regulations to promote design-for-environment and green manufacturing methodologies for new electrical and electronic products. Firms will most probably pursue such strategies, in order to reduce the burdens associated with regulation compliance [3].

2.2 Take-Back Systems

As the responsibility for the disposal of products at the end of their useful life shifts back to the producers, companies have been forced to incorporate these take-back requirements into their strategic planning processes and product design strategies [6, 7, 8, 9]. Through a take-back system industrial products are recovered at the end of their useful life, so that waste is diverted away from landfills and incineration into reuse, remanufacture and recycling, as shown in figure 1.

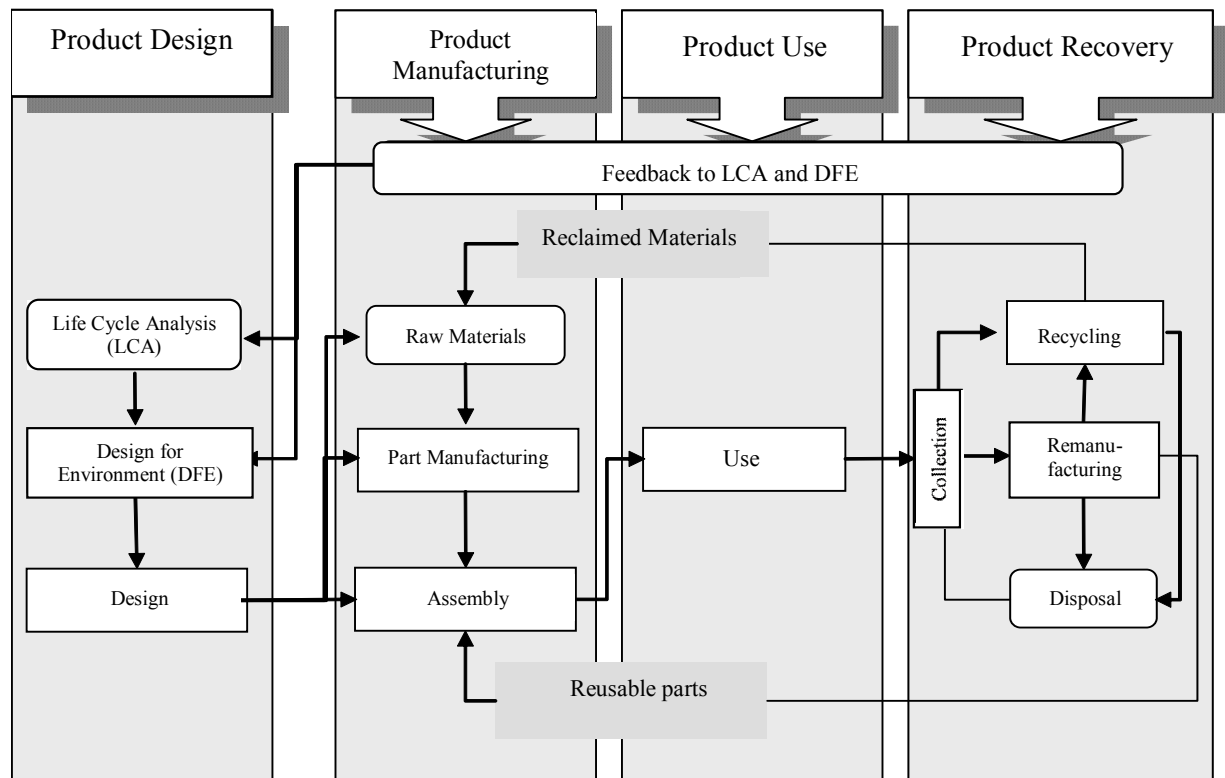


Figure 1 Interactions among the activities in a product life cycle (Source: [10]).

The paramount objective in the practice of industrial ecology and sustainable manufacturing is to create an industrial ecosystem in which all discarded products become the raw material inputs for new products. In an idyllic closed-loop industrial ecosystem, the raw material input requirements of the manufacturing processes are exactly balanced by the waste supplied from the discarded products. The motivation behind reusing waste as an input for new production is environmentally and ecologically pure, and the initiative for undertaking an industrial ecology approach represents sound policy. However, the enforced introduction of these well intentioned industrial ecology practices can subsequently create very negative, unintended environmental consequences, particularly for electronic products [3].

Apart from legislation, there are many other significant incentives that motivate companies take the responsibility of managing end-of-life products. These drivers can be environmental, economic and also marketing, since the corporation's image plays a significant role to consumers' choice [7, 11, 12, 13].

As far as environmental drivers are concerned:

- There are landfill bans for specific products.
- The capacity of landfills has been reduced dramatically.
- Environmental accountability forces companies to use reusable items and materials.

The following drivers can be recognized as economic:

- Used products represent cheap resources from which value can be recovered.
- Recovery is often cheaper than building or buying new products of virgin materials.
- Recovering used products may limit dependence on suppliers. Product recovery may also act as an extra source of supplies when a supplier is unable to fulfil an order.
- Product recovery may also take less time than purchasing or producing new products. This may result in shorter delivery times as well as lower storage costs.
- A manufacturer that takes back and resells refurbished products creates an additional source of income, since the unit margin of each product increases.
- Instead of taking advantage of the increased marginal profit, the manufacturer can partially pass these benefits on to the customer through price discounts. Consequently, he would gain from selling the used product on the secondary market, while also increasing the market share.

Finally, as far as marketing drivers are concerned:

- Used product take back and recovery is an important element for building a green profile based on corporate citizenship and improving the company's marketing position. "Green" products that meet other requirements can have competitive edge, since consumers are willing to pay higher prices to protect the environment.
- Potential competition between virgin products and recovered products can be partly avoided. If a company does not collect and reuse its products itself, someone else might and this may have adverse effects on the companies brand. Companies may also take back products to prevent sensitive components from leaking to secondary markets.

In order to formulate and implement the appropriate take-back system, the infrastructure of the supply chain, reverse flows and processes / activities that take place must be taken into account. Typically reverse supply chains comprehend the following five groups of activities, which are linked by intermediate transportation [11]:

- *Collection:* All activities rendering used items (product, component or material) available and physically moving them to some point for further treatment. This may involve product acquisition, transportation and storage.
- *Inspection & Separation:* Results in splitting the flow for various recovery and disposal options. This may involve testing, disassembly, shredding, testing, sorting and storage.
- *Re-processing:* Reusable flows undergo the actual transformation of a used item into a reusable item of some kind. Depending on the recovery option chosen, this comprehends various activities such as disassembly, shredding, repair, replacements etc. The available recovery options are the following:
 - Direct reuse
 - Repair
 - Refurbishing
 - Remanufacturing
 - Cannibalisation
 - Scrap
- *Disposal:* The non-reusable flows are disposed of to incinerators and landfills.
- *Re-distribution:* Directing reusable items to a market to new markets, and physically moving them to potential new users. This involves sales activities, transportation and storage.

As a result, the manufacturer has to a) decide whether he will take advantage of the forward supply chain flows or not, b) look into the possibilities of making the reverse flows more stable, so that the collection of end-of-life products is accomplished (first activity of the take-back process), and c) look into the alternative choices as far as the management of end-of-life products is concerned (second to fifth activity).

If reverse flows can be handled through the same structure as forward flows, investments regarding the reverse supply chain infrastructure are likely to be low. There are however many reasons for not to combine reverse and forward flows: different volumes, different handling requirements, different timing, and different destination. Moreover, truck loading and unloading procedures are more complex, if at each stop material is being unloaded and loaded. Furthermore, the point of delivery and collection are not necessarily the same at each location [7]. However, the manufacturer should locate the recycling facilities close to customer markets in order to ease the direct delivery of used products from end-users [14].

In any case the company should also look into possibilities of making the reverse flows more stable to influence supply quantity [15, 16]. Tools that can be used for this purpose are presented below [17].

- *Deposit fee*: Deposit fees are paid when an item is bought and then reimbursed when it is returned according to specifications. The disadvantage of deposit fees is that they make buying the product more expensive.
- *Buy-back option*: At the moment the product is sold, the buyer is offered the possibility to return the product to the seller for a preset price, provided it fulfils some preset requirements (e.g. use, timing).
- *Reduced price “new”*: Buyers get a discount from buying a new version, when they return the old version.
- *Pay a fee*: Buyers pay a fee when they return the product for recovery. It basically works like a disposal fee.
- *Take back for free*: Buyers do not have to pay when they return the product for recovery.
- *Gift*: Every time a product is returned for recovery a non-profit organization receives an amount of money.

The main decision a company faces regarding the management of end-of-life products is whether to be active or passive. An active alternative is one where the company decides to collect end-of-use products and dispose of them, resell them as is or reprocess and redistribute them (see the thick axis line in figure 4). The alternative is passive if the manufacturer limits his involvement to paying, letting a third party (recycling company) take care of returned products (see the discontinuous line in figure 4). However, in any of the two alternatives the activities of the take-back process remain the same. The only difference lies in the entity that takes care of the activities, as shown in figure 2.

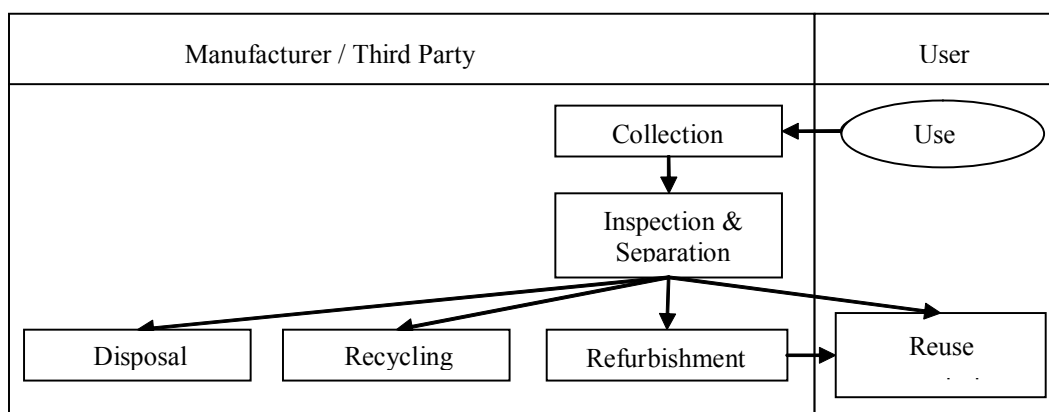


Figure 2 Take-back process activities.

In figures 3 and 4 an indicative forward supply chain and an indicative reverse supply chain are depicted, respectively. It is obvious that the collection activity takes place in more than one echelon. Specifically, in the reverse supply chain of figure 4, end-users take back the products at the “initial collection stations”, which lie at a local level. Then, these products are transferred and stored at the “intermediary collection stations”, which lie at a regional level. Next, the products get transferred and stored at the two “central warehouses”, which may lie in the north and south region of the country, respectively. From these central warehouses the returned products are finally collected from either the manufacturing company or a third recycling party that is supposed to manage them.

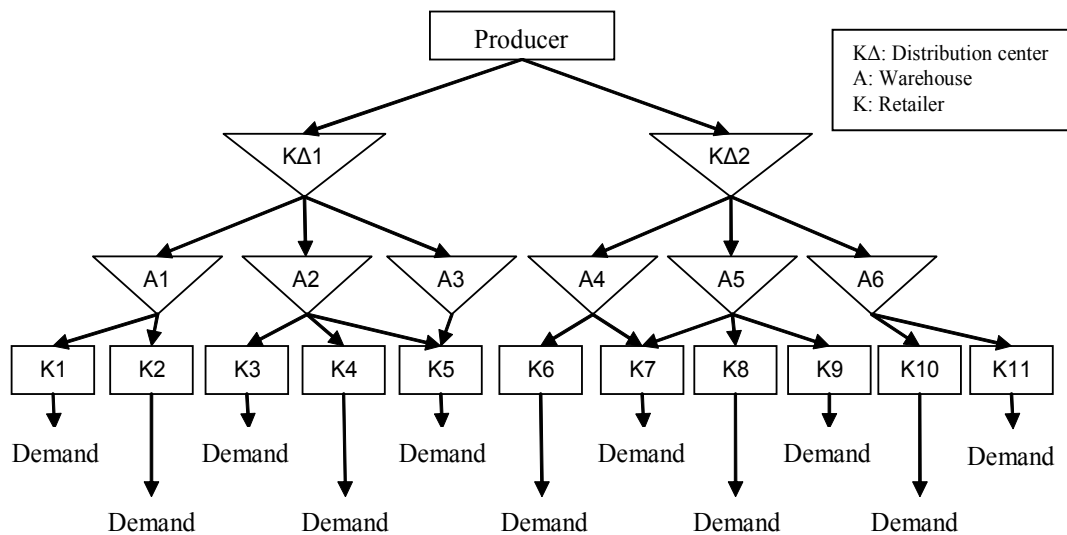


Figure 3 Indicative forward supply chains.

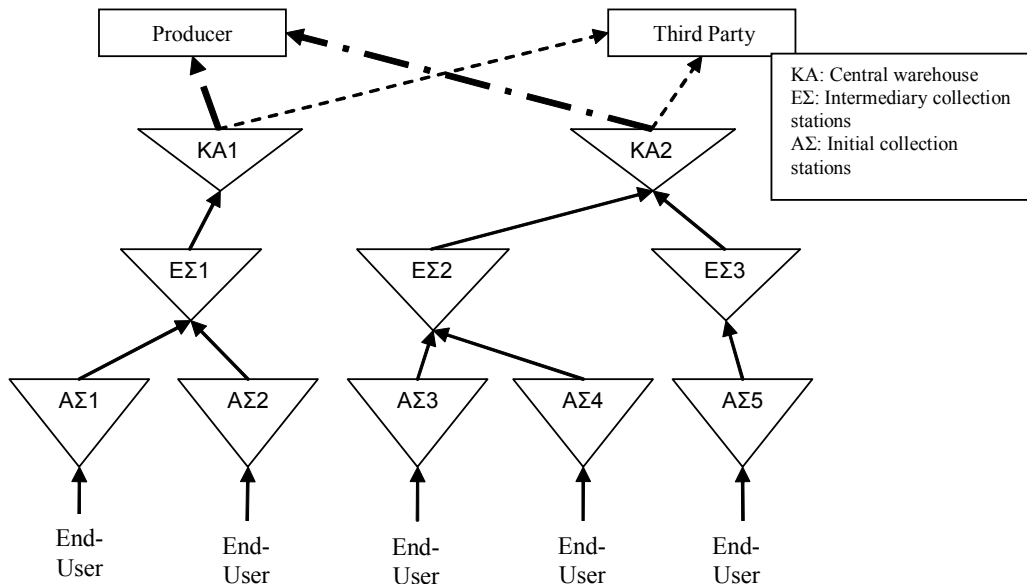


Figure 4 Indicative reverse supply chain.

Finally, it should be mentioned that take-back infrastructure is especially well developed in the Netherlands, and other countries are expected to develop similar programs in the near future [6]. Apart from the development of the appropriate take-back system, the manufacturing company can formulate

and implement an integrated environmental management system. Such a decision depends on the corporate strategy, which is determined by the internal and external environment of the company. Consequently, the producer has to assess the dynamic environment to take the right decision. The manufacturer must also identify his customers' requirements (for example, the ISO 14001 certification) and the benefits that can be obtained from the implementation of an environmental management system, in order to choose the right alternative.

2.3 The ISO 14001 Standard

In 1996 International Organization for Standardization (ISO) developed the ISO 14000 series, which describe the requirements to be fulfilled by organizations to implement an effective environmental management system (EMS). An effective environmental management system can be integrated with other managerial functions to assist organizations in achieving environmental and economic goals [1].

The ISO 14000 series consists of 21 standards and guidance documents, based on a plan-do-check-act framework. These environmental standards, as shown in figure 5, are divided into six categories: 1) life cycle assessment, 2) design for environment, 3) environmental labels and declarations, 4) environmental communication, 5) environmental performance evaluation, and 6) environmental management systems auditing.

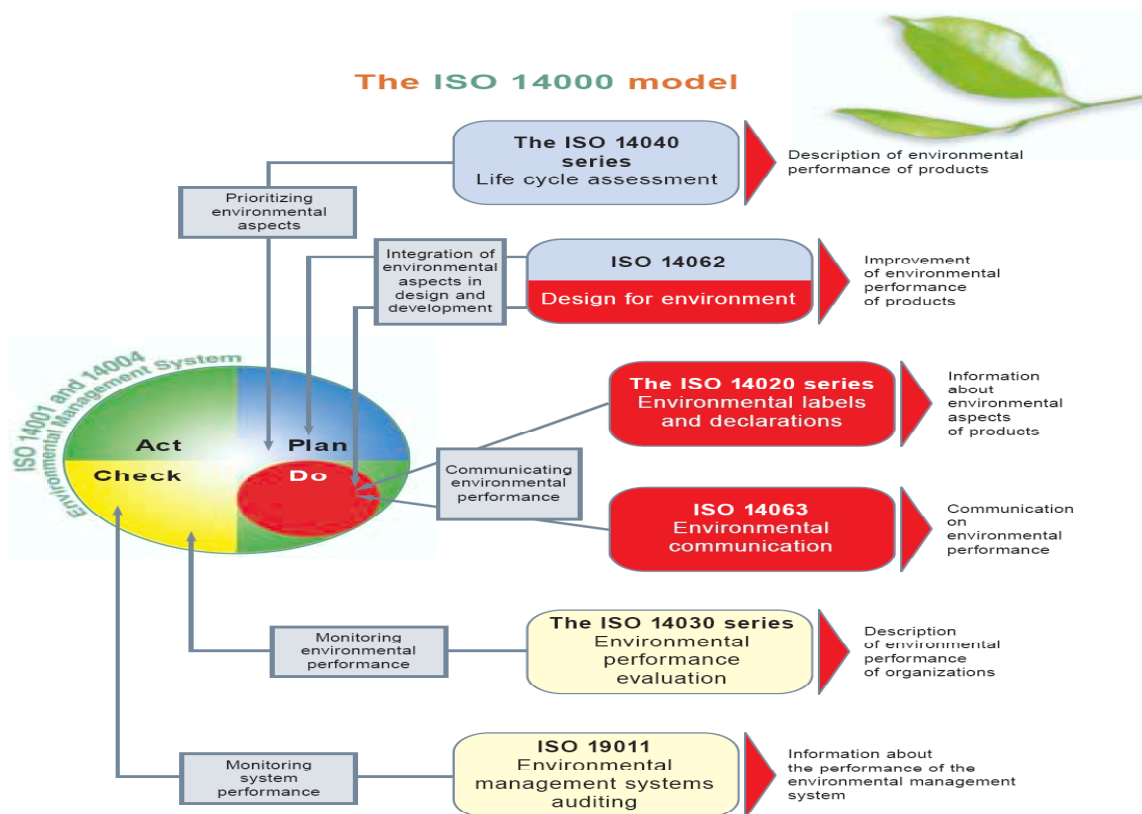


Figure 5 The ISO 14000 model (Source: [18])

In 2004 the ISO 14001 standard was updated to the ISO 14001:2004 standard. ISO 14001 is considered the only standard designed for the purpose of audit and certification in the ISO 14000 series [19]. The core elements of the ISO 14001:2004 standard are environmental policy, planning, implementation and operation of programs to meet objectives and targets, checking and corrective action, review and improvement [18].

In Greece, ELOT (Hellenic Organization for Standardization) is the unique certification organization, authorized for the granting of the Hellenic Marks of Conformity. ELOT has developed and operates a scheme for the certification of Environmental Management Systems, according to the requirements of

the ISO 14000 series of standards and the EA (European Accreditation) Guidelines for bodies certifying Environmental Management Systems [20]. The certificate ELOT awards is called “Environmental Management System Certificate of Conformity ELOT EN ISO 14001”.

Today, an increasing number of corporations globally are certifying their environmental management systems by ISO 14000 series standards. Advocates of ISO 14001 claim substantial operational, managerial, and competitive benefits for corporations that adopt the international guidelines. On the other side, critics contend that ISO 14001 does not ensure either legal compliance or continued performance improvements. They claim that at plants or facilities already complying with environmental regulations, ISO 14001 certification may merely be an image-building or public relations effort [21].

According to Rondinelli and Vastag [21], ISO 14001 could theoretically serve as a framework for significantly improving performance in a firm with minimal environmental management capacity, i.e. a *panacea*, in a sense, or as a set of *common sense* guidelines for enhancing performance in a firm with regulatory compliant practices. At the same time, some firms may simply use ISO 14001 as a ‘*label*’ for green image building.

Some firms are using ISO 14000 guidelines to develop new environmental management systems, or adapt their environmental practices to the international standard, without formally certifying them. Other corporations, government agencies, and environmental interest groups are sceptical about the real impacts of ISO 14000 certification, and either ignore the guidelines or question their effectiveness in improving environmental performance. In essence, the ISO 14001 certification acknowledges that a company has an acceptable environmental management program [21].

According to Babakri et al. [22], it takes most companies between 8 to 19 months to obtain the certificate. Moreover, Babakri et al. [19] conducted a survey to study the impact of the ISO 14001 certification on recycling performance of bottling companies. According to their survey, 1) the recycling performance of a company is significantly positively affected by certification, 2) the recycling performance improvements due to the ISO 14001 certification are significantly lower for larger companies than it is for smaller ones, and 3) earlier certified organizations experience better recycling performance than more recently certified organizations. However, according to Rondinelli and Vastag [21], the ISO 14001 guidelines simply assume that good environmental management systems will, if they are implemented effectively, reduce or eliminate negative environmental impacts and move a company toward better environmental performance.

ISO 14001 has proven to be a useful tool to evolve from maintaining regulatory compliance to a position of improved productivity and enhanced competitive advantage. There is mounting evidence that companies which manage not only the standard economic factors but also the environmental and social factors affecting their business show financial performance superior to those which fail to manage all three [18].

ISO 14000 registration may benefit organizations by helping them to: 1) comply with a major customer, who may require registration to ISO 14000, 2) achieve a competitive advantage in the global market, 3) improve compliance with applicable environment laws, regulations, and standards, 4) reduce their exposure to environmental liability, 5) increase community goodwill by preventing pollution and waste, and 6) reduce the number of audits required by regulatory agencies [1].

Furthermore, ISO 14001 is a framework that inspires and channels the creativity of all members of an organization, making them active agents of change promoting environmental protection, resource conservation and improved efficiencies. When all members in an organization are challenged to think differently, it leads to the creation of innovative products and services [18].

Finally, the significant tangible economic benefits obtained by ISO 14000 registration are: 1) reduced raw material and resource use, 2) reduced energy consumption, 3) improved process efficiency, 4) reduced waste generation and disposal costs, and 5) utilization of recoverable resources [20].

2.4 Environmental Management Systems

As we already stated in the previous subsection, the ISO 14000 series is the starting point for the formulation and implementation of an environmental management system. The requirements of the series determine the necessary steps for the implementation of an effective EMS.

The prime focus of an environmental management system is to prevent adverse environmental effects and improve environmental performance by institutionalizing various environmental programs and practices, such as initiating environment related performance measures and developing green technologies, processes and products [23]. An EMS involves integrating environmental thinking into new product development, including design, material selection, manufacturing processes, and delivery of the product to the consumers, plus the end-of-life management of the product after its useful life [10].

According to Gupta [23], an environmental management system should:

- identify and assess the environmental effects arising from the organization's existing or proposed activities, products or services
- identify and assess the environmental effects arising from incidents, accidents and potential emergency situations
- identify the relevant regulatory requirements
- enable priorities to be identified and pertinent environmental objectives and targets to be set
- facilitate planning and control, auditing and reviewing activities to ensure both that the policy is complied with, and that it remains relevant
- be capable of evolution to suit changing circumstances.

World-class companies are continuously improving their operations, in order to outperform compliance with the minimum requirements and become pioneers in their industry. However, according to a survey conducted by Gutowski et al. [6] in the automotive and electronic sector, the overwhelming majority of the companies (>90%) are struggling to balance business goals and environmental goals.

It is obvious that management of end-of-life products, which can be achieved through a take-back process, represents a subset of an environmental management system. Undoubtedly, the take-back process requires that the products can be disassembled, so that the appropriate components can be recovered. The producer's role is important in this phase, since he has to decide whether we will implement a take-back system to comply with legislation or he will formulate and implement an integrated environmental management system taking advantage of its benefits.

The reasons behind promoting reverse logistics practices are of both an economic as well as environmental kind. Among the economic motives we find the recovery of the value still incorporated in the used product and the important savings in materials and components. From the environmental viewpoint, we might cite concern regarding solid waste pollution, landfill saturation or the scarcity of raw materials [24].

Competitive Economic Advantage	Proactive Green Behaviour	Regulatory Mandates
Reduced waste treatment and disposal costs	Corporate image	Banned materials and reporting requirements
Conservation of energy, water, materials	Regulatory flexibility	
Reduced compliance costs	Employee satisfaction	Product take-back requirements (EU, Japan)
Supply chain requirements	ISO 14001 Certification	
First to achieve product compliance	Market value of company	Emissions standards
First to achieve cost-effective product take-back system	Green purchasing Eco-labeling	Worker exposure standards

Table 1 Motivating factors and actions for EMS (Source: [6]).

According to Gutowski et al. [6], the factors and actions that motivate companies adopt environmental management systems are shown in table 1.

3. Motivating the Active Involvement of Consumers

Consumers have the crucial voice and the entire system revolves around their attitudes. It is they, who support products through their purchasing behaviour. Additionally, consumers control the final disposition of the products, ultimately choosing to responsibly dispose of the product through the indicated channels or to irresponsibly dispose of the product in the trash bin or unauthorized dump [8]. Hence, consumers must be motivated to participate actively in the take-back processes, otherwise these systems will collapse [25].

According to Stevels [26], there are seven distinct types of consumers when it comes to environmental issues:

- Environmentally Engaged (E.E.)
- Environmental Optimists (E.O.)
- Disoriented Consumers (D.C.)
- Environment too Complicated (E.C.)
- Environmental Pessimists (E.P.)
- Growth Optimists (G.O.)
- Enjoy Life (E.L.)

This list shows a wide diversity of consumer attitudes that can lead to a diversity of behaviour with respect to environmental issues. Therefore, environmental consciousness is required from consumers, while, on the other side, companies and governments must motivate people by means of education, and information as far as the benefits from the contribution to environmental management are concerned.

Green consumerism is today at all-time high levels. The consumers are willing to pay premiums for goods, which are recycled, recyclable and non-damaging to the environment. In the USA 80 per cent of shoppers polled said that protecting the environment was so important that it warrants any cost [23, 27]. According to Gutowski et al. [6], in Europe there is a very high level of public awareness of environmental issues that has propagated up into the government often through elected “Green Party” officials. Furthermore, according to Tsoufas and Pappis [14], consumers are environmentally conscious, since they realize their contribution to environment protection and are attracted to companies that have invested in green technologies.

According to Straughan and Roberts [28], environmental commitment and behaviour is associated with demographic and psychographic characteristics of consumers. These characteristics along with their interaction with ecological consciousness are following. The first five characteristics are demographic, while the next three psychographic.

- *Age*: Younger individuals are likely to be more sensitive to environmental issues.
- *Sex*: Women are more likely than men to hold attitudes consistent with the green movement. Theoretical justification for this comes from the fact that women will, as a result of social development and sex role differences, more carefully consider the impact of their actions on others.
- *Income*: Income is generally thought to be positively related to environmental sensitivity. The most common justification for this belief is that individuals can, at higher income levels, bear the marginal increase in costs associated with supporting green causes and favoring green product offerings.

- *Education*: Education is not positively correlated with environmental concerns and behaviour.
- *Place of residence*: Those living in urban areas are likely to show more favorable attitudes towards environmental issues.
- *Political orientation*: Those with more liberal political beliefs are more likely to exhibit strong verbal commitment than those with more conservative political views.
- *Perceived consumer effectiveness*: Consumers' attitudes and responses to environmental appeals are a function of their belief that individuals can positively influence the outcome to such problems.
- *Environmental concern*: There is a positive correlation between environmental concern and environmentally friendly behaviour.

As mentioned above, the typical profile given for a green consumer is a young, mid- to high-income, educated, urban woman. Moreover, it is apparent that, if consumers are convinced that they can contribute to the environment protection, then it is highly possible that they will adopt a green behaviour. Furthermore, it was proven above that, when people are aware that their actions pollute the environment, they believe that their contribution can solve this problem. Therefore, it is necessary to motivate consumers through advertising campaigns, mass media informing, environmental education, reinforcement of the environmental movement, and communication of the green message.

As far as consumer motivation is concerned, it can be achieved through the take-back system. The tools used for the stabilization of the reverse flows in the supply chain can be rather helpful in this case. According to Brisson [25], consumers must be given incentives to generate less waste, whether it will be through recycling and greater reuse of packaging, or by demanding products which are packed in less packaging from the manufacturers.

Through a deposit fee, a reduced price new or a take back for free scheme the consumer is pushed to take an active part in the take-back system and return the products. Moreover, when the pay a fee scheme refers only to non-recyclable parts or to components that cannot be recovered, this forces consumers indirectly to prefer green products, in order to reduce the cost they face from waste generation.

Consumers resent when they find difficulties in the disposal of products at the end of their useful life or even when they have to pay for the disposal. The possibility to irresponsibly dispose of the product in the trash bin or unauthorized dump is rather high. In such cases, the pay a fee scheme only for the non-recyclable parts and the components that cannot be recovered, along with a reduced price new scheme for a green product can motivate consumers adopt a green behaviour.

As far as consumer motivation is concerned, corporations, organizations and governments have to promote take-back systems through advertising campaigns, in order to inform people about environmental issues, green products, and green behaviour. In this way, they will contribute to the reinforcement of the environmental movement and they will communicate the green message. Furthermore, a long-term formulation is necessary, regarding the environmental education, so that people finally acquire environmental culture.

By means of advertisement, companies can affect consumers' choice and clarify the perceived consumer quality. The issue of perceived quality is very important, and firms employ different tactics to convince customers that they are not purchasing a product that is inferior in quality. The most effective tactic is a lower price combined with an "as if new" warranty [16]. It is obvious that price plays the most significant role in consumers' motivation and is a means to convince them that they do not sacrifice quality by purchasing a green product.

4. Summary & Conclusions

During the past twenty years there has been a growing interest in environmental concerns by the public, government, and business community. Environmental laws and regulations have been enacted

and also corporations have taken voluntary initiatives to demonstrate their commitment to environmental consciousness and obligations. This environmental consciousness brought in a growing interest in the flows of the reverse supply chain, where consumers return the goods back to the producers at the end of their useful life, while producers adopt take-back systems in order to collect and manage these products. Usually, manufacturers choose to take up environmental management systems, because a voluntary initiative, like the ISO 14001 certification, is part of their corporate environmental policy, and also plethora of benefits and incentives can be obtained from the implementation of such systems.

As the responsibility for the disposal of end-of-life products shifts back to the producer, companies are forced to incorporate these take-back requirements into their strategic planning processes. Therefore, the manufacturer's critical role in the reverse supply chain is to recognize the legislation and formulate the appropriate reverse supply chain, deciding whether he will manage the disposed products or outsource them to a third party. Moreover, the producer has to identify his customer's requests (for example the ISO 14001 certification) and finally the benefits that can be obtained from the implementation of an environmental management system. In other words, the producer has to choose between compliance with the minimum requirements (implementation of a take-back system) and implementation of an integrated environmental management system (proactive green behaviour).

Through a take-back system industrial products are recovered at the end of their useful life, so that waste is diverted away from landfills and incineration into reuse, remanufacture and recycling. In an idyllic closed-loop industrial ecosystem, the raw material input requirements of the manufacturing processes are exactly balanced by the waste supplied from the discarded products. For the implementation of a take-back system, the manufacturer has to a) decide whether he will take advantage of the forward supply chain flows or not, b) look into the possibilities of making the reverse flows more stable, so that the collection of end-of-life products is accomplished, and c) look into the alternative choices as far as the management of end-of-life products is concerned.

Today, an increasing number of corporations are certifying their environmental management systems by ISO 14000 series standards. Advocates of ISO 14001 claim substantial benefits for companies that adopt the guidelines, while critics contend that the certification 14001 does not ensure either legal compliance or continued performance improvements. In essence, the ISO 14001 certification acknowledges that a company has an acceptable environmental management program. The ISO 14000 series is usually the starting point for the formulation and implementation of an environmental management system. The requirements of the series determine the necessary steps for the implementation of an effective EMS.

Consumers have the crucial voice and the entire system revolves around their attitudes. They support products through their purchasing behaviour and they control the final disposition of products. Hence, they must be motivated to participate actively in the take-back processes. The typical profile given for a green consumer is a young, mid- to high-income, educated, urban woman.

Consumer motivation to generate less waste and dispose of the products responsibly can be achieved through the take-back system. The tools used for the stabilization of the reverse flows in the supply chain can be rather helpful in this case. As far as consumer motivation is concerned, corporations, organizations and governments have to promote take-back systems through advertising campaigns, in order to inform people about environmental issues, green products, and green behaviour. They must also reinforce the environmental movement and communicate the green message. Furthermore, a long-term formulation is necessary, regarding the environmental education, so that people acquire environmental culture.

In this paper we identified the environmental laws and regulations regarding industrial electrical and electronic products, we summarized the key points regarding the environmentally related ISO 14001 certification and the implementation of either a take-back or an environmental management system, and finally we captured the role of manufacturers in the reverse supply chain along with methods to motivate customers participate in product take-back systems. This theoretical study is certainly not exhausted. The take-back system presented in this paper is indicative and needs to be further analyzed and specialized to different scenarios, in order to become comprehensive.

Acknowledgement

Sponsored by the Greek Ministry of National Education and Religious Affairs for the project: “Pythagoras: Reinforcement of research groups in universities” and under the title: “Optimum Management of Industrial Products at the End of their useful Life (MIPEL)”.

References

- 1 Rezaee Z, Elam R. Emerging ISO 14000 Environmental Standards: A Step-By-Step Implementation Guide. MAJ 2000; 15 (1/2): 60-7.
- 2 Murphy P. A Preliminary Study of Transportation and Warehousing Aspects of Reverse Distribution. TJ 1986; 25: 12-21.
- 3 Halluite J, Linton D, Yeomans J S, Yoogalingam R. The Challenge of Hazardous Waste Management in a Sustainable Environment: Insights from Electronic Recovery Laws. CSREM 2005; 12: 31-7.
- 4 European Union. Directive 2002/96/EC of the European Parliament and of the Council of 27 January 2003 on Waste Electrical and Electronic Equipment (WEEE). OJEU 2003; L 037: 24-38.
- 5 European Union. Directive 2002/95/EC of the European Parliament and of the Council of 27 January 2003 on the Restriction of the Use of Certain Hazardous Substances in Electrical and Electronic Equipment. OJEU 2003; L 037: 19-23.
- 6 Gutowski T, Murphy C, Allen D, Bauer D, Bras B, Piwonka T, Sheng P, Sutherland J, Thurston D, Wolff E. Environmentally Benign Manufacturing: Observations from Japan, Europe and the United States. JCP 2005; 13: 1-17.
- 7 Herold M. Managing End-Of-Use Products. Licentiate Dissertation 2004, Helsinki University of Technology, Finland.
- 8 Rose C M, Stevels A. Applying Environmental Value Chain Analysis to Product Take-Back Systems. Proceedings of the 7th CIRP International Seminar on Life Cycle Engineering, 27-29 November 2000, Tokyo, Japan.
- 9 Stevels A L N, Ram A A P, Deckers E. Short Report: Take-Back of Discarded Consumer Electronic Products from the Perspective of the Producer; Conditions for Success. JCP 1999; 7: 383-9.
- 10 Gungor A, Gupta S M. Issues in Environmentally Conscious Manufacturing and Product Recovery: A Survey. CIE 1999; 36: 811-853.
- 11 Krikke H R, Le Blanc H M, Van De Velde S. Creating Value from Returns. CentER Applied Research Working Paper No. 2003-02. January 2003.
- 12 Jacobsson N. Emerging Product Strategies: Selling Services of Remanufactured Products. Licentiate Dissertation 2000, Lund University, Sweden.
- 13 Heese H S, Cattani K, Ferrer G, Gilland W, Roth A V. Production, Manufacturing and Logistics: Competitive Advantage Through Take-Back of Used Products. EJOR 2005; 164: 143-157.
- 14 Tsoulfas G T, Pappis C P. Environmental Principles Applicable to Supply Chains Design and Operation. JCP 2005; Article in Press.
- 15 Guide D, van Wassenhove L. Business Aspects of Closed Loop Supply Chains. Carnegie Mellon University Press, Pittsburgh, 2003.
- 16 Ayres R, Ferrer G, van Leynseele T. Eco-Efficiency, Asset Recovery and Remanufacturing. EMJ 1997; 15 (5): 557-574.
- 17 Flapper S. Product Recovery Strategies in Guide and van Wassenhove 2003, Business Aspects of Closed Loop Supply Chains. Carnegie Mellon University Press, Pittsburgh, 2003.
- 18 International Organization for Standardization. ISO 14000 <http://www.iso.org>, accessed: 30 May 2006.
- 19 Babakri K A, Bennett R A, Rao S, Franchetti M. Recycling Performance of Firms Before and After Adoption of the ISO 14001 Standard. JCP 2004; 12: 633-7.
- 20 Hellenic Organization for Standardization. Systems & Products Certification <http://www.elot.gr>, accessed: 30 May 2006.

- 21 Rondinelli D, Vastag G. Panacea, Common Sense, or Just a Label? The Value of ISO 14001 Environmental Management Systems. EMJ 2000; 18 (5): 499-510.
- 22 Babakri K A, Bennett R A, Franchetti M. Critical Factors for Implementing ISO 14001 Standard in United States Industrial Companies. JCP 2003; 11: 749-752.
- 23 Gupta M C. Environmental Management and its Impact on the Operations Function. IJOPM 1995; 15 (8): 34-51.
- 24 González-Torre P L, Adenso-Díaz B, Artiba H. Environmental and Reverse Logistics Policies in European Bottling and Packaging Firms. IJPE 2003; 88: 95-104.
- 25 Brisson I. Packaging Waste and the Environment: Economics and Policy. RCR 1993; 8: 183-292.
- 26 Stevels A. Green Marketing of Consumer Electronics. Proceedings of Electronics Goes Green Conference, 11-13 September 2000, Berlin, Germany.
- 27 Denton D K. Enviro-management as a Profit Center. EMH 1998; 9/2: 60-4.

Alignment of Core Ideology and Organizational Values with Strategy and Tactics

Vassilis Kafatos¹, Leslie Szamosi², John Kawalek³

¹SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, vkafatos@seerc.info

²City College, Dept. of Business Administration and Economics, 2 Kalapothaki Str, 54624 Thessaloniki, Greece, lszamosi@city.academic.gr

³The University of Sheffield, Management School, 9 Mappin Str, Sheffield, S1 4DT, UK, j.kawalek@shef.ac.uk

Organizational values are a key component of strategic planning and a defining element of company culture. Based on consulting experience with multiple companies, it is believed that continuous alignment of these values with strategy and tactics can substantially enhance overall performance. The research that is in progress will try to identify the key attributes of how the core purpose and values of corporations are successfully translated into strategies and operational practices and how can alignment among them be sustained in world of constant change. High performing companies will be identified and analyzed based on a value creation framework and case studies will be used to test the validity of the resulting concept. The findings are expected to have significant benefits in overall corporate performance. This paper outlines the purpose and objectives of the research, provides a brief overview of the relevant strategic planning literature and its association with organizational values, sets the principles of the research method s to be used and discusses the potential benefits of the anticipated outcomes.

Keywords

Alignment, strategy, strategic planning, organizational values.

1. Introduction

The goal of the research is to identify the key attributes of how values are successfully translated into strategies, tactics and operational practices.

In more detail the research will try to answer the following questions:

- What are the best practices through which the core ideology (values and purpose) of the corporation is embedded in its strategy, tactics and operations?
- How can alignment of values with strategy, commitment and capability be assessed and sustained in a world of constant change?
- Is there a significant difference in the attributes of translating values into strategy and operations between companies in countries of different culture and philosophies (e.g. Greece compared to Southeast Europe)?

The research findings are expected to be of high practical value in enhancing performance, through the development of a comprehensive framework that can guide corporations in translating their core ideology and values into strategic decisions and tactical initiatives and monitoring ongoing alignment in a world of constant change. Moreover, as more companies are expected to have operations throughout different countries in South East Europe, the importance of enhancing the understanding of the role that varying and even changing cultural norms play in this process of translation of values into strategy, operational practices and behaviours, becomes even more important.

2. Background

The development of the field of strategic management has been dramatic over the last decade, in an effort to respond to increased level of globalization, the rise of the information age, constant change and uncertainty and lately to increased emphasis on ethics. Recently, re-emerging is a research trend towards linking strategic planning with organizational values (Williams, S.L., 2002, [1]) while experience has shown the importance of practicing values such as trust and openness to the achievement of strategic imperatives and overall corporate performance (Deloitte, 2005). Consulting experience suggest that in most companies stated values do not actually become organization reality and practice, as strategic planning is rarely linked with values, leadership fails to exercise ownership and practice the value statement, and the corresponding behaviours are not fostered through management practices, processes and systems, with negative impact on performance and growth.

Several management researchers in their quest to discover the keys to high performance have identified core values as a basic focus area (Kirby, 2005, [2]).

Among them, Peters and Waterman (1982, [3]) identified eight principles that characterize excellent companies. They have found that excellent companies are value driven, have a “well-defined set of guiding beliefs” and do their best to support their values into their organizational culture.

Collins and Porras (1994, [4]) have found that the essence of a visionary company comes in the translation of its core ideology, which consists of core values and core purpose, and its own unique drive for progress, into every element of the organization (goals, strategies, tactics, policies, processes, cultural practices, management behaviors, etc.). Moreover they have introduced the concept of alignment, in the sense that all the elements of the company should work together in concert within the context of the company's core ideology and its vision. According to their research, visionary companies preserve a cherished core ideology while stimulating progress and change in their business strategies and operating practices.

In the same line, Lencioni (2002, [5]) stressed the importance of living by authentic stated corporate values that need to be integrated into every employee-related process. In that sense, sharing a common set of corporate values within an organization, determines what are worthwhile activities, ethical behaviours and moral responsibilities, has a strong impact on strategic direction (De Wit, 2004, [6]) and can contribute to sustained performance and value creation over the long term.

Investigating further the issue of alignment, Kaplan and Norton (1996, [7]) have developed the Balanced Scorecard as a framework to clarify and translate vision and strategy and link strategic objectives with measures. In order to translate strategy into operational terms they have identified four key performance areas (financial, customer, internal business process and learning and growth) that a company should focus on in order to succeed. The Strategy Map developed later (2004, [8]), provided a graphic tool that assists in the alignment of organizational strategy with the company's intangible assets.

Over the last years, many researchers have emphasized internal resources, capabilities, assets and competencies (e.g. resource-based view, strategic leadership, knowledge based view) as being primary drivers of competitive advantage (Hoskisson, Hitt, Wan, & Yiu, 1999, [9]). Eisenhardt and Sull (2001, [10]) have argued that in rapidly changing and ambiguous markets, strategy implementation can not be based neither to a unique position with tightly integrated activity system, nor to valuable and inimitable resources, but rather on focusing on key strategic processes and developing simple rules. In other words, when business becomes complicated, strategy should be simple. Being true to core values and ideology and embedding those to strategy and tactics can capitalize on simplicity, while preserving the flexibility to pursue the company's vision and goals.

With global expansion and heightened public awareness of ethical issues, strategists need to understand the influential variables and dominant moral philosophies, in order to establish organization code of ethics and corporate and translate them in strategies and actions (Roberson, C.J., & Crittenden, W.F., 2003, [11]). Along the same line, the proposed research will also explore how society's ethical norms and values, that are dependent on the dominant moral philosophy of the

particular country, can affect the process of incorporating corporate core values to strategies and operating practices.

3. Method

The case methodology, based on archival and interview data and in-depth case studies, will be used, as it can provide much richer information about the firm's idiosyncrasies and capture the intangible nature of the research issues (Hoskisson, Hitt, Wan, & Yiu, 1999, [9]). Moreover, according to Feuer & Chaharbaghi (1995, [12]) as the level of dynamics in the environment increases, the research should focus on developing conceptual knowledge by heavily relying on qualitative studies.

During the research, companies will be identified from a range of industries that outperformed their peers in generating shareholder value. The cases that will emerge as top performers will be analyzed based on extensive examination of public information about them (books, newspaper and magazine articles, financial statements, corporate annual report, web site documents, etc.) and interviews with executives and consultants with expertise on these companies.

To ensure systematic and comprehensive data collection and analysis, the Enterprise Value Map (Deloitte, 2004) will be employed (Figure 1). The Enterprise Value Map framework, developed by Deloitte, provides a linkage among Shareholders Value, Strategy and Tactics and a comprehensive tool of looking at a company. It starts from Shareholder Value, identifies the Value Drivers (i.e. how value is created) and structures the improvement levels in terms of strategy (change what you do) and tactics (do what you do better). By using the Enterprise Value Map as a framework for reviewing the selected companies, we will examine how their core ideology and values, either written or implicit, are or can be integrated into strategy and tactics and what factors enable and sustain alignment among them.

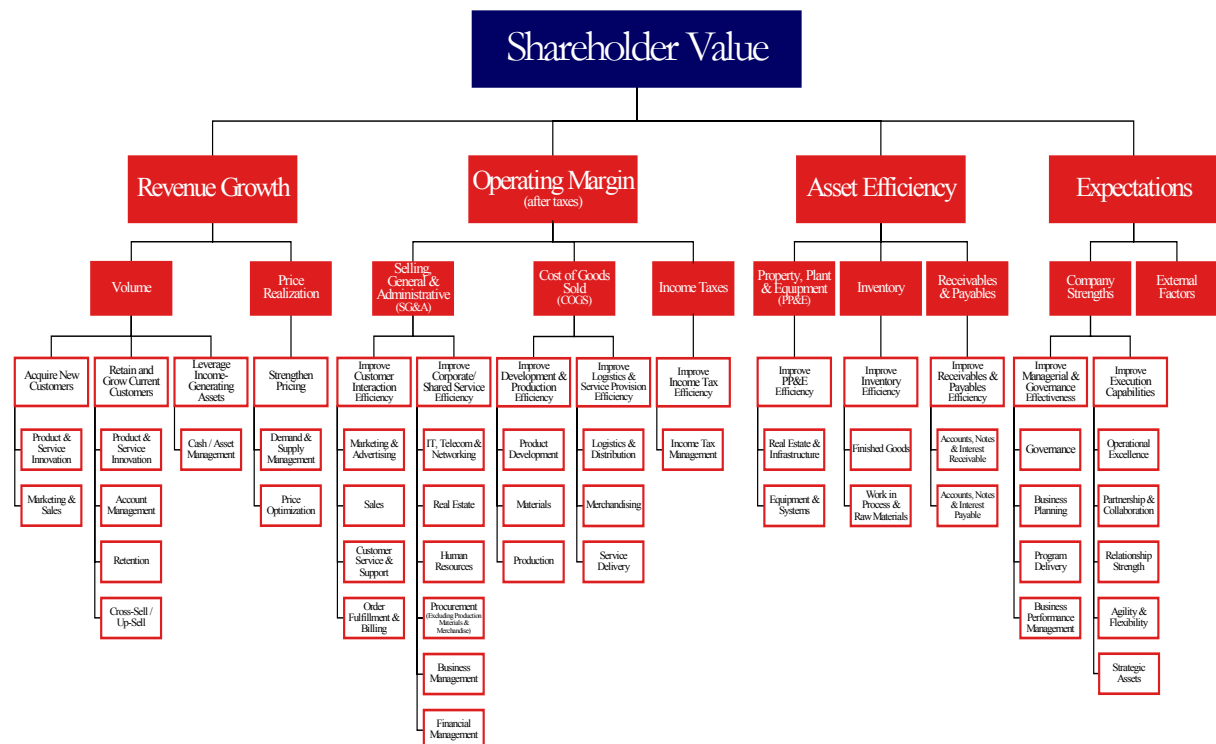


Figure 1 Deloitte Enterprise Value Map TM (condensed version).

Next, the findings and concepts will be tested in real world, via consulting engagements or detailed case studies in companies in Greece and Southeast Europe. This process will provide a valuable feedback loop that will help to improve our concept as we progress through the research.

4. Discussion

The research will capitalize on the researcher's work and management consulting experience as Partner with Deloitte & Touche Consulting and on extensive business databases and access to numerous companies, through which the finding's validity can be enhanced.

To enhance alignment of strategic planning and organizational values Williams (2002, [1]) discussed a human resources development process that includes assessment or audit of current value beliefs, value statement setting, training and communicating value-based practices, communicating value-based performance and fostering value-centered environment. Trickett (1997, [13]) has developed a values audit, in order to provide a method for encouraging honest discussion of values and their linkage to strategies, purpose and performance.

The research results will hopefully enhance understanding and provide a useful management tool on how to achieve and monitor alignment between corporate core values and strategic targets, practices and behaviors, in order to close the gap between ideology, strategy and reality and enhance shareholder value in the long term. In more detail the research hopes to develop a managerial tool for monitoring that core values are directly linked with corporate and business strategy (strategy), provide guidance for the desired culture, leadership and talent within the organization (commitment) and are incorporated in organizational, managerial and operational processes (capability). In today's society, shared and practiced core values may prove to be the unifying parameter that enhances consistent focus towards its purpose and strategic direction and longevity in the marketplace.

References

- 1 Williams, S.L. Strategic planning and organizational values: links to alignment. Human Resource Development International, 2002; 5:2, 217-233
- 2 Kirbi, J. Towards a Theory of High Performance. Harvard Business Review, 2005; 7-8
- 3 Peters, T., & Waterman, R.H. In Search of Excellence: Lessons from America's Best-Run Companies. Harper & Row. 1982
- 4 Collins, J.C., & Porras, J.I. Built to Last: successful habits of visionary companies (1st ed.). HarperBusiness, New York. 1994
- 5 Lencioni, P.M. Make your values mean something. Harvard Business Review, 2002; 7: 113-117.
- 6 De Wit, B. & Meyer, R. Strategy process, content, context (3rd ed.). Thomson Learning. 2004
- 7 Kaplan, R.S., & Norton, D.P. The Balanced Scorecard. Harvard Business School Press. 1996
- 8 Kaplan, R.S., & Norton, D.P. Strategy Maps. Harvard Business School Press. 2004
- 9 Hoskisson, R.E., Hitt, M.A., Wan, W.P., & Yiu, D. Theory and research in strategic management: Swings of a pendulum. Journal of Management, 1999; 3: 417-456.
- 10 Eisenhardt, K.M., & Sull, D.N. Strategy as simple rules. Harvard Business Review, 2001; 1: 107-116.
- 11 Roberson, C.J., & Crittenden, W.F. Mapping Moral Philosophies: Strategic Implications for Multinational Firms. Strategic Management Journal, 2003; 24: 385-392.
- 12 Feurer, R & Chaharbaghi, K. Researching strategy formulation and implementation in dynamic environments. Benchmarking for Quality Management & Technology, 1995; 4: 15-26.
- 13 Trickett, D. How to Use a Values Audit. Training & Development, 1997; 51(3):34-8.

Cluster Policy and Entrepreneurship Promotion in Transition Countries in SEE: The Case of Bulgaria, Republic of Macedonia (FYROM) and Serbia

Aleksandar Karaev¹, S.C. Lenny Koh², Leslie T. Szamosi³

¹SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, akaraev@mt.net.mk

²The University of Sheffield, Management School, 9 Mappin Street, Sheffield, S1 4DT, UK, S.C.L.Koh@sheffield.ac.uk

³City College, Dept. of Business Administration and Economics, 2 Kalapothaki Str, 54624, Thessaloniki, Greece, szamosi@city.academic.gr

In the globalization process factors such as flexibility, speed and innovation contributed entrepreneurship to emerge as a key player in driving economic development. There is also an abundance of literature to support the contribution of clusters in strengthening the entrepreneurship and increasing the competitiveness of Small and Medium Enterprises. The literature also suggests that competitiveness is not only a derivation of entrepreneurial activity of individual firms, but also a result of an appropriate governmental cluster policy. This is one of the reasons why many developing countries instead of providing sector-based government interventions, which in some cases lead to market distortion, are embracing cluster based policies and entrepreneurship promotion strategies as a potential instrument for accelerating economic development. Besides establishing basic macroeconomic framework conditions for economic growth, the governments in transition economies can play an important role in active facilitation of cluster development processes and supporting entrepreneurial initiatives. This is a literature review paper, which aims to give an overview of cluster policies and entrepreneurship promotion strategies in the selected countries in South East Europe.

Keywords

Clusters, entrepreneurship, SMEs.

1. Introduction

In order to respond to the challenges of the globalization process macro level governmental policy tries to improve the competitiveness of the national economies through creating favourable framework conditions for economic activity and promoting various instruments for SME development. According to the literature based on experience in some industrialized countries, clusters contribute toward improving the performance of SMEs, and increasing the competitiveness of certain geographical regions and even nations themselves. Recognizing their potential positive role and needs for assistance, many governments have been active in designing and implementing policies and incentives to support SME development through both financial and non-financial instruments. In a same time the existence of an environment favorable for business start ups and strong entrepreneurial spirit has been widely recognized as one of the main preconditions for creation and development of clusters.

After discussing the literature on cluster theory, this literature review paper examines general concept of cluster policy. This is followed by an overview of policies for entrepreneurship promotion and development of clusters in selected transition economies as a base for further research in this field.

2. Cluster definition

Over the last decade, clusters have been widely recognized as one of the ways of overcoming the size limitations of companies and as an important instrument for improving their productivity, innovativeness and overall competitiveness. Despite the fact that numerous studies have been conducted in various countries, a common understanding of the cluster concept has yet to be achieved. According to Porter, widely considered to be one of the most prominent authorities in the field, national clusters are formed by firms and industries linked through vertical (buyer/supplier) and/or horizontal (common customers, technology etc.) relationships with the main players located in a single nation/state [1]. This definition was later expanded by including institutions (formal organizations) such as universities [2]. Accordingly, geographical proximity has been seen as a conduit towards facilitating the transmission of knowledge and the development of institutions, which in turn may enhance cluster effectiveness. According to Porter's views, clustering can encourage an enhanced division of labour among firms with physical proximity among numerous competing producers, thereby encouraging innovation.

Other authors support that clusters refer to geographically bounded concentrations of interdependent firms, which should have active channels for business transactions, dialogue and communication [3]. Without active channels, even a critical mass of related firms is not generally perceived as a local production or social system, and therefore does not operate as a cluster per se [3]. Clusters consist of private enterprises of various sizes, including producers, suppliers, and customers, plus labor, government, professional associations, and academic, research or training institutes.

The United Nations International Development Organization (UNIDO) applies a cluster definition related to the sectoral and geographical concentrations of enterprises that produce and sell a range of related or complementary products and, face common challenges and opportunities. These concentrations give rise to external economies such as the emergence of specialized suppliers of raw materials and components or growth of a pool of sector-specific skills and can foster development of specialized services in technical, managerial and financial matters [4].

In Japan two types of clusters can be distinguished; first, the so-called *jiba-sangyo* (or localized industrial communities of the traditional type), where SMEs link to each other as industrial clusters and second, the geographically concentrated *sangyo-shuseki* (or industrial agglomerations in a particular locality) of a more recent origin, where SMEs gather together in support of each other in a new industrial activity, around a large-sized enterprise as input suppliers, or around an academic community (universities and research institutions) [5]. The main difference between industrial districts and the cluster approach is that the first is more input oriented, securing geographically available inputs for production, and the second is based on generating optimal competitive conditions for firms [6].

It is difficult to precisely determine which factors are prerequisites for cluster development and which are the result of the clustering process. The geographical proximity of markets and suppliers, the existence of a pool of specialized labour, the presence of input equipment, the availability of specific natural resources and infrastructure, low transaction costs due to geographic proximity among actors and access to information have been commonly cited as requirements for the creation for cluster creation [7]. Clusters have also been identified as forming naturally form as a result of a perceived common interest of its members and the stakeholders are SMEs, business associations, local and regional governments, business service providers and supporting institutions and each participant in the process of cluster development needs to identify, articulate and realize its own role [7].

One of the factors that complicate comparisons between clusters is their varying geographical coverage. Some regional clusters are greater in size and population than national clusters in smaller countries. A *cluster's* boundaries depend mainly on the linkages between cluster participants and complementarities across industries and institutions that are most important to competition [2]. Cluster boundaries do not have to comply with political ones and can cover a cross-border area.

Successful examples of cluster approaches are offered both from regions focusing on "traditional products", such as furniture, ceramics and food (Northern Italy), and from regions with predominantly

high technological product outputs (Silicon Valley). Cluster based economic development has proven highly successful, in both smaller and larger EU countries. Many successful case studies indicate that the coordination of economic activities – depending on the intensity of cooperation in the form of clusters – can also strengthen the competitiveness of, in particular, national economies. A good example is the so-called “The Chair triangle” in the Udine Region and Friuli Venezia Giulia in Northern Italy, which produces 80% of total Italian chair product and 50% of total European production. It covers an area of 100km², on which 1.200 companies are concentrated with 15.000 employees and annual turnover of 2,5 billion Euro [8]

3. Cluster policy and successful cluster approach

Policy makers have identified industrial clusters as potential engines for economic growth and innovation. Cluster policy is not only an industrial policy but also a socio-cultural one. Policy makers should determine the place of the cluster policy with regard to the overall economic policy of the country. This is particularly important since considerable financial support needs to be allocated to the projects, and the capacity of each country to do so varies significantly [9].

A specific industrial policy is considered as a cluster policy if it is oriented to the promotion of specific regional characteristics and if it aims, in a structural sense, to make a contribution to the further development of branch concentration or network building blocks for clusters or to the further development of existing clusters [10].

A cluster policy should provide a framework for dialogue and cooperation between firms, the public sector (particularly at local and regional levels of government) and non-governmental organisations [9]. In general, cluster policy can be implemented in one of two ways: (1) assisting the development of an existing and already established cluster, and (2) creating a new cluster deploying external knowledge and experience [2]. Porter (1998) further underlines that government should not create clusters artificially, when there are no preconditions for that, but should reinforce and build on already established and emerging clusters as was evident in the previous examples of Southern Italy. Both approaches to cluster policy share some specific characteristics: focus on local systems or regions instead of on individual companies, promotion of SMEs instead of large companies, reliance on internal strengths, promotion of social capital as an important factor of cluster development - encouraging trust-based relationship to increase the flow of knowledge between local players rather than intervening, for example, through financial incentives [11].

According to some authors, competitiveness means that individual firms, the firms of a sector or the firms of a region or of a country can successfully assert themselves in the domestic market and on the world market [12]. It was also suggested by the same authors that competitiveness is not only a result of entrepreneurial activity of individual firms, but also a result of an appropriate structural policy, functioning competitive policy and adequate infrastructure. Therefore, the effectiveness of collaboration between firms, related institutions and government structures can strongly influence the performance of a particular cluster.

4. SME sector and cluster development in selected transition countries

4.1 Introduction

SME sector and cluster concept have been widely researched in the developed countries. SME sector can be seen a backbone of EU economy, since 99 % of all enterprises in the EU is small and medium [13]. “Statistics on SMEs in other countries indicate a significant contribution made by SMEs to the total employment, turnover and GDP” (p.7) [13].

There is an abundance of literature and history of practical experience in enterprises organising in clusters in industrialised countries; however, since this literature review paper intends to review the cluster based strategies in the selected transition countries, this chapter will give an overview of

support of the SME sectors, governmental cluster policies in these countries as well as cluster initiatives from international donor organizations.

4.2 Cluster policy and initiatives in SMEs in the Republic of Macedonia (FYROM)

The importance of SMEs in the Republic of Macedonia (FYROM) is evident by the fact that in 2004 they represented 99 % of total number of active enterprises in the country [13].

According the National SME Strategy until 2003, the macroeconomic policy, as well as other working documents of the Government, the development of SMEs in Macedonian economy has high priority [14]. For realising this aim, the Government has prepared the following documents to support SME development: Strategy for Development of SME, Act on SME Development Support, Law on Macedonian Guarantee Agency, Law on Realisation of Handicraft Activities, National Policy for Development of Technology.

All these documents have an objective to facilitate the establishment of a favourable institutional, legislative, administrative and financial environment for the establishment and development of SMEs in the country.

The country adopted the European Charter for small enterprises on the Thessaloniki's Summit in June 2003 and since then, every year the Ministry of Economy elaborates national report for the current situation of small and medium enterprise in the country. "By becoming a member of the European Charter for small enterprises, Republic of Macedonia (FYROM) has decided to work towards the realization of the constant improvement in the areas defined in the Charter and to cooperate with other member countries in order to reach the stated objectives, through the process of benchmarking of growth and exchange of positive experiences and practices" (p.1) [15].

The following public and private institutions are set up to regulate, develop and promote entrepreneurship [16]:

National Council for Competitiveness and Entrepreneurship provides a platform for dialog between the Government and the private sector and it was established with goal of improving the business environment in the country [15].

Entrepreneurship Promotion Department within the Ministry of Economy (SMEs Department) is responsible for creation of policies, measures and projects towards SMEs [15].

Agency for Promotion of Entrepreneurship (APE) is a governmental institution, established in May 2004 to implement the Program on measures and activities for promotion of entrepreneurship and creation of competitiveness of SMEs (the National SME strategy) as a successor of the National Enterprise Promotion Agency [15].

Besides the official governmental institutions, the SME sector has been supported by technical assistance from various donor organizations. The main characteristic of the programs for technical assistance is that they are supporting SMEs through capacity building of business associations and supporting the transfer of knowledge among SMEs as well as between SMEs and public sector and educational institutions.

In the Republic of Macedonia (FYROM), activities have already been undertaken by the international donor organizations for a financial, organizational, infrastructure, scientific and consultant support of the process for the establishment and development of clusters. The following donor organizations implement projects, in cooperation with the SME department of the Ministry of Economy, which underlined the importance of the cluster policy in the Program for supporting entrepreneurship [14].

USAID's Competitiveness Activity - The mission of USAID's Competitiveness Activity (MCA) is to build prosperity by helping enterprises become stronger and more competitive in the global marketplace. One of its main goals is to promote and strengthen industry clusters that will develop and execute strategies for improving the international competitiveness of SMEs. Clusters consist of private enterprises of various sizes, including producers, suppliers, and customers, plus labour, government,

professional associations, and academic, research or training institutes [17]. MCA has identified and supports the following clusters:

- *Lamb and cheese*: Entrepreneurs in the lamb and cheese cluster are promoting the export of their products through establishing strong links between sheep herders, dairies, slaughterhouses, training and education institutions, government agencies and ministries. With over seventy members they are trying to create new economic opportunities for the cluster. Given the long tradition of producing sheep cheese and lamb products, the cluster hopes to increase the income of its members through is elaborating a joint export strategy, built on cooperative linkages among companies and related associations [17].
- *Tourism*: The tourism cluster consists of more than sixty representatives from travel agencies, business associations, educational organisations and public institutions. This cluster tries to increase the turnover in the tourism sector by attracting high income visitors, who would be interested to learn and to explore not only natural beauties, but would also look for receiving a broader cultural experience [17].
- *Information Technology (IT)*: The IT provides a platform through which members can become more acquainted with industry practices, learn about their mutual strengths and propose joint projects. The cluster works with the Government on improving policy conditions for the IT industry, attract foreign direct investment, adapt the educational system to meet IT challenges and creating the appropriate legal framework for e-commerce [17].
- *Wine*: The wine cluster utilizes country's unique varieties to increase premium production and export while developing a country wine brand. The members of the wine cluster are wineries, grape growers' associations, research institutes, wine production equipment suppliers and donor organizations, who want to approach together domestic and foreign markets [17].
- *Apparel*: The apparel cluster is focusing on producing value added products, instead exporting cheap labor only, improving the flexibility of own production, restoring the local production of fabrics, developing Macedonian proprietary designs and brands, and attracting foreign direct investments [17].

GTZ Private Sector Promotion Project supported the cluster development process indirectly through creating preconditions for cluster development in the four selected industrial sectors: Information Technology, food processing, tourism and wine. The project focuses on strengthening the links between the private sector, the Government, business service providers, educational institutions, financial resources and international organizations, through organizing training programs, joint trade fair participations, capacity building of business associations, etc [18].

The Italian Institute for Foreign Trade (ICE) supports development of shoe and leather cluster in the North Eastern part of the country [19].

4.3 Support of the SME sector and cluster policy in Serbia

The SME sector in the Republic of Serbia has been constantly growing over the last ten years. The data from the year 2000 show that in Serbia there are 60.552 active enterprises, and the majority are small enterprises – 94,1% (56.993), then medium enterprises - 4,3% (2.573), and big enterprises only 1,6% (986) [20].

The Serbian Government has set two priority targets for development of SME sector by the end of 2007 [21]:

- Increase the total number of SMEs from 270.000 to 400.000, and
- Creation of over one million new jobs in the SME sector.

Institutions for supporting SME development in Republic of Serbia are the following:

- the Ministry of Economy and Privatization, having two departments: Department for SME development and Department for Private Entrepreneurship – with a major task of leading the creation of a simulative business environment for SME sector in the Republic of Serbia,
- Republic Agency for Development of SMEs and Entrepreneurship - responsible for implementing SME strategy, promoting SME sector and coordinating activities of Regional SME agencies
- Regional agencies and Centres for Development of SMEs and Entrepreneurship, which have a role as a catalyst and implementers for practical support services to enterprises [21].
- SME Advisory Board (SMEAB) Advocacy for SMEs – Established in 2001, the SME Advisory Board (SMEAB) is a national Forum for advocacy for SMES and promotion of public-private dialogue [22].

Within a framework of activities for improving competitiveness the “Proposal for Measures for Export Promotion” has been developed, endorsed by the Serbian Government in May 2004. As part of competitiveness development policy, a draft “Cluster Policy” has been prepared and an initiative for cluster mapping has been started. Currently the process of creating automotive and wine clusters is underway and activities aiming at creating preconditions for establishing ICT cluster have been started [23].

Besides the governmental institutions for supporting SME sector and entrepreneurship, there is a strong presence of international organizations:

German Organization for Technical Cooperation (GTZ) is implementing the program “Promotion of economic development and employment in Serbia”, on behalf of the German Ministry of Economic Cooperation and Development, and Ministry of Economy and Privatization of Serbia. The program has been implemented in two basic fields: regional and local economic development and support to economic development in five industries (textile and leather, fruits and vegetable production, automotive industry, IT – software and tourism) [24].

European Union implements programs through the European Agency for Reconstruction (EAR) [25].

United States Agency for International Development (USAID) launched a significant competitiveness activity 2003, which is focused on industry clusters [26]. The competitiveness building activities are implemented through the Serbian Enterprise Development Project (SEDP), which aims at increasing exports in a two target industry clusters: fruit and fruit juice producers, and furniture design and manufacturing. Each one of these includes 20 to 25 companies, which received technical assistance in management, financial operations, marketing, production organization, introduction of new technologies, equipment, etc. In the process of identifying additional clusters, there is an on-going assessment of possibilities in the field of clinical testing, pharmaceuticals, textile industry and tourism [26].

4.4 SME sector and cluster policy in Bulgaria

Besides the fact that SMEs represented 99,8 % of the total number of private enterprises in Bulgaria in 2003 (p.20), “they also generate 79 % of employment, 75 % of the turnover, 61 % of the value added of private enterprises, 54 % of the export and 73 % of the import of private business” [27]. This proves the well-known conclusion that private firms are generally better performing than state-owned ones and are really the most efficient segment of the economy [28]. The influence of the SME sector to the Bulgarian overall economic growth has been emphasized in the National Strategy for Encouragement of Small and Medium Enterprises Development for 2002-2006, which was accepted by the Bulgarian Government. In the Strategy “the main strategic objective is to create a favorable environment and conditions for development of a competitive SME sector, which will accelerate the economic growth of the country through preservation of macro-economic stability.” (p.3) [29]

In order to meet the strategic objective, the strategy defines the main priorities for addressing challenges in the SME sector. Some of those priorities have indirect influence on the cluster

development policy and others are directly mentioning clusters as a tool for achieving the strategy goals.

Bulgaria introduced the European Charter for Small Enterprises (approved by EU leaders at the Feira European Council on 19-20 June 2000) on 23 April 2002 in Maribor, Slovenia [30]. According to the EU Charter for Small Enterprises, small enterprises must be considered as a main driver for innovation, employment as well as social and local integration and therefore the best possible environment for small business and entrepreneurship needs to be created [30]. Entrepreneurship as one of the basis for clusters development is integrated in both the main principles of the Charter and in the priorities of the National Strategy. "The government will encourage the entrepreneurship as a valuable and productive characteristic" (p.2) [29]. Besides recognizing and rewarding successful enterprises, the EU Charter for Small Enterprises considers failure only as a learning opportunity related to initiative and risk taking behaviour [30].

Along the same line is the article 14 from the Law for small and medium enterprises [31] which stresses "creating a positive image of entrepreneurship through programmes and measures for its promotion and supporting research of opportunities to implement new business ideas" (p.2)

Under the strengthening the technological capacities of small enterprises, the Charter aims to promote the relationship between SMEs, through the formation of clusters.[30]. In 2004 the Council for Economic Growth discussed and approved the draft of the Bulgarian Innovation Strategy. One of the proposals in the Strategy is the development of clusters in Bulgaria. The Draft Innovation Strategy was adopted by the Council of Ministers [32].

The main institution for supporting SME development in Bulgaria is Bulgaria's SME's Promotion Agency, which facilitates SME development through, encouraging start-ups, promotion of SME Development and increasing the SME competitiveness. The main priority of the Agency is to implement the National SME Promotion Strategy [33].

Besides national support, the Bulgarian SME sector receives support from multilateral and bilateral donor organizations. "Currently the EU with nearly \$ 350 million per year is the largest multilateral donor in Bulgaria, with most of its assistance aiming to prepare Bulgarian accession into EU. World Bank assistance to Bulgaria has been set at \$750 million for 2002 - 2005. The United Nations Development Program centers its Country Cooperation Framework on good governance, job creation, information technology, and environmental protection" (p.2) [34]. In the field of SME and cluster support the most active bilateral donors in Bulgaria are: USAID, and Swiss German and Japanese Governments. Not all of them are directly involved in cluster development, but their activities, such as institutional development, management training, start-up support are directly or indirectly supporting SME development. They coordinate their activities and their timing and priorities are in line with the major EU accession requirements.

The largest bilateral donor in Bulgaria in the field of SME development, USAID, has funded The Bulgaria Competitiveness Exercise, which is being implemented by Management Systems International in association with J.E. Austin Associates. After analysing different industry sectors, BCE supported industry leaders from the tourism, canning and food, textiles and garments, information technology, wine, and transport industries to begin the process of developing industry competitiveness strategies [35]. Although five clusters were originally selected, only ICT received long-term project assistance [36].

5. Findings

In all of the selected countries, SMEs are recognised as a main engine of economic growth, and therefore they have developed different policies for stimulating the development entrepreneurship and SME sector. Following the experiences of developed countries, the Republic of Macedonia (FYROM), Serbia and Bulgaria, have accepted the cluster approach, to varying degrees, as an instrument for improving their national competitiveness and have integrated the cluster policy in their main documents for SME development. The official governmental policies for SME development are

additionally supported by international organisations such as USAID, UNDP, GTZ, ICHE, and others which provide both, financial and technical assistance. As main designers of cluster support projects, they are selecting industrial sectors for starting cluster initiatives and setting measures for attracting the potential cluster members. Unfortunately, in some cases the selected sectors are not complementary with the national strategies for economic development and there is a lack of coordination between the governmental programs and those of the international organisations. In addition, international organisations offer different models of supporting cluster development and propose different strategies for implementing measures, which creates confusion among SMEs in their overall perception of clusters. They also have different methods for impact monitoring and evaluation of results of their cluster initiatives. Following the desire to present better results to the main funding organisations of cluster initiatives, in some cases the positive effects of clusters are overemphasized. In reality there is no strong evidence that cluster policy brings additional positive effect to the existing SME policy in the transition countries. Such effects have not been researched especially from the point of view of the SMEs, the main actors in the cluster development process, that their performance has been improved as a result of cluster effects.

6. Conclusions and future research

According to Schwanitz et al (2002), competitiveness means that individual firms, the firms of a sector or the firms of a region or of a country can successfully assert themselves in the domestic market and on the world market [12]. It was also suggested by the same authors that competitiveness is a result of entrepreneurial activity of individual firms, but also a result of an appropriate structural policy, functioning competitive policy and adequate infrastructure. Therefore, the effectiveness of collaboration between firms, related institutions and government structures can strongly influence the performance of a particular cluster.

Governments - both national and local - have considerable roles to play in the promotion of a clustering approach. They should actively promote such an approach, besides creating the framework conditions, setting the rules for competition and promoting entrepreneurial spirit [2].

All cluster participants need assistance in this process of strengthening their levels of trust, cooperation, and in the development of effective private – public dialogue. In the process of role definition, international donor organizations should find their place in providing support measures on all levels, taking into consideration the sustainable development of the country in a sense that economic benefits are available for everyone [37]. There are some indicators, however, that clusters with high levels of dependence on foreign assistance are less autonomous, have weaker capabilities and have difficulties in achieving long-term sustainability [38].

The aim of the study is to give an overview of cluster policies and entrepreneurship promotion strategies in the selected countries in South East Europe. In the phase of the further research the types of results that cluster-oriented assistance strategies bring in the transition countries in SEE will be examined. The research will focus on how SMEs perceive cluster policy and entrepreneurship promotion programs offered by governmental institutions. Besides the governmental policy support of cluster development, international donor organizations also implement projects aimed at transferring experiences from countries where clusters contribute to economic development. Hence, special attention will be paid to investigate the experience of SMEs with cluster support projects offered by international donor organisations. For this, the methodology used will be a different regional survey of SMEs, members of different clusters as well as governmental and international cluster supporting organizations.

Telephone and personal interviews will also be adopted for qualitative data collection. Overall, it can be noted that this research attempts to use both quantitative and qualitative methods.

References

- 1 Porter, M., 1990, *The Competitive Advantage of Nations*. New York.

- 2 Porter, M., 1998, Clusters and the New Economy of Competition. Harvard Business Review, Nov/Dec98, Vol. 76 Issue 6, pp.77-91.
- 3 Rosenfeld, S.A., 1997, Bringing Business Clusters into the Mainstream of Economic Development. European Planning Studies, 5, pp.3-23.
- 4 UNIDO, 2000, Promoting Enterprise through Networked Regional Development. UNIDO publications
- 5 Ozawa, T., 2003, Economic Growth, Structural Transformation, and Industrial Clusters: Theoretical Implications of Japan's Postwar Experience. Proceedings of the Conference on Clusters, Industrial Districts and Firms: The Challenge of Globalization held at University of Modena and Reggio Emilia. Modena, University of Modena and Reggio Emilia.
- 6 Preissl, B. and Solimene, L., 2003, Innovation clusters: Virtual links and globalization. Proceedings of the Conference on Clusters, Industrial Districts and Firms: The Challenge of Globalization held at University of Modena and Reggio Emilia. Modena, University of Modena and Reggio Emilia.
- 7 Gallo, C. and Moehring, J., 2002, Innovation and clusters, Proceedings of the East-West Cluster Conference OECD LEED. Grado, Italy.
- 8 Organization for Economic Cooperation and Development, 1997, OECD Globalisation and SME, Synthesis report.
- 9 Andersson, T., Serger, S. S., Soervik, J. and Hansson, W.E. (2004) The Cluster Policies Whitebook. IKED, Malme, Sweden.
- 10 Bruch-Krumbein, W. and Hochmuth, E. (2000) Cluster und Clusterpolitik – Begriffliche Grundlagen und empirische Fallbeispiele aus Ostdeutschland, Verlag Schüren, Marburg. Cited in: GTZ (2002) Cluster study.
- 11 Boekholt, P. and Thuriaux, B. (1999) Overview of cluster policies in International perspective. A report for the Dutch Ministry of Economic Affairs. Cited in: "Regional clusters in Europe ". Observatory of European SMEs 2002/No 3.
- 12 Schwanitz, et al., 2002, Competitiveness of Economic Sectors in EU Accession Countries: Cluster-Oriented Assistance Strategies. Deutsche Gesellschaft Fur Technische Zusammenarbeit (GTZ), Eschborn.
- 13 EAR - European Agency for Reconstruction (2005) Macedonian SME Observatory Report 2004. EAR, October 2005, p.1-31.
- 14 MoE (2003) Ministry of Economy of Republic Macedonia (FYROM) Program for promoting entrepreneurship and competitiveness of SMEs in Republic Macedonia, Skopje, MOE.
- 15 MoE (2005) Ministry of Economy of Republic of Macedonia (FYROM), European Charter for Small Enterprises, National report 2005.
- 16 German Organization for Technical Cooperation, GTZ (2001) - Private Sector Promotion (GTZ-PSP), Business Service Institutions in Republic of Macedonia (FYROM), Final Report, Web Site [online] Available from: http://www.gtzipsp.com.mk/Uploads/wwwInstAnal_FinReport.doc Accessed [20th February 2006].
- 17 USAID/ Project Macedonian Competitiveness Activity – MCA (2004) MCA Web Site [online]. Available from: www.mca.org.mk [Accessed 30th March. 2006]
- 18 GTZ – PSP (2003) German Organization for Technical Cooperation, GTZ (2001) - Private Sector Promotion (GTZ-PSP), Web Site [online]. Available from: <http://www.gtzipsp.com.mk> Accessed [20th February 2006].
- 19 Italian Institute for Foreign Trade ICE (2006) Web Site [online]. Available from: <http://www.italtrade.com/countries/europe/macedonia/sto.htm> Accessed [20th February 2006].
- 20 SADSME - Serbian Agency for Development of SMEs and Entrepreneurship, (2003). Available from: <http://www.sme.sr.gov.yu/Deo2Eng2003.pdf> [Accessed 19th Sept. 2005]
- 21 Ministry of Economy and Privatisation of Serbia and Montenegro (2003). A Strategy for the Development of Small and Medium-sized Enterprises and Entrepreneurship in the Republic of Serbia, Belgrade.
- 22 D'Aniello, L. (2003) Policies for business in the Mediterranean Countries: Serbia and Montenegro, Centre for Administrative Innovation in the Euro-Mediterranean Region
- 23 Ministry of Internal Economic Relations of Serbia, (2004) Work Results in 2004 Web site [online] Available from <http://www.mier.sr.gov.yu/upload/dokumenta/WORK%20RESULTS2004.pdf>

- 24 German Organization for Technical Cooperation, Serbia – GTZ (2005) Available from National Association for local authorities (<http://www.skgo.org/code/navigate.php?Id=169>) [Accessed 18th Sept. 2005]
- 25 EAR - European Agency for Reconstruction (2006), Serbia Available from http://www.ear.eu.int/publications/main/documents/Serbia_2003_12.pdf [Accessed 30th March 2006]
- 26 USAID/ Serbian Enterprise Development Project – SEDP (2005) USAID Web Site [online]. Available from: http://www.usaid.org.yu/center/program_news/programs.php?id=123®ion=s&parent_id=2 [Accessed 18th Sept. 2005]
- 27 Center for Economic Development (2004) Annual Report on the Condition and Development of Small and Medium-sized Enterprises in Bulgaria' 2004, prepared under assignment of the minister of economy and energy. [online]. Sofia, Ministry of Economy and Energy of Republic of Bulgaria. Available from: <http://www.mi.government.bg/eng/ind/econ/docs.html?id=144158>). Accessed [20th December 2005].
- 28 Alexandrova, S., Blagoeva Z. Bogdanov, L., Chavdarov, V., Dimitrov, L., Dimitrov, T., Ganev, G., Gochev, A., Kopeva, D., Kotupova, M., Rozenov, R., Popov, A., Stanchev, K. (1998) Creating competitive clusters in Bulgaria: factors, vision and strategies. Joint Report of the Institute for Market Economics and the World Bank Institute.
- 29 Ministry of Economy and Energy of Republic of Bulgaria (2002) National Strategy for Encouragement of Small and Medium Enterprises Development for 2002-2006. [online]. Sofia, Ministry of Economy and Energy of Republic of Bulgaria. Available from: <http://www.mi.government.bg/eng/ind/econ/docs.html?id=144158>). Accessed [20th December 2005].
- 30 European Charter for Small Enterprise (2000) Feira European Council, June 2000 Available from http://europa.eu.int/comm/enterprise/enterprise_policy/charter/index_en.htm [Accessed 29th March 2006]
- 31 Ministry of Economy and Energy of Republic of Bulgaria (2001) Law for Small and Medium Enterprises Development [online]. Sofia, Ministry of Economy and Energy of Republic of Bulgaria. Available from: <http://www.mi.government.bg/eng/ind/econ/docs.html?id=143590> Accessed [20th December 2005].
- 32 European Commission, Enterprise Directorate- General (2005) European Trend Chart of Innovation, Annual Innovation Policy Trends and Appraisal Report, Bulgaria
- 33 Bulgarian SME's Promotion Agency (2004) SME Promotion Agency Web Site [online]. Available from: <http://www.sme.government.bg/en/default.asp>. Accessed [23rd November 2005].
- 34 USAID (2006) Web Site [online]. Available from: <http://www.usaid.gov/policy/budget/cbj2005/ee/bg.html> [Accessed 30th March. 2006]
- 35 MSI The Bulgarian Competitiveness Initiative (2001) MSI Web Site [online]. Available from: <http://www.competitiveness.bg/project.htm>. Accessed [10th of December 2005].
- 36 Silcox, S., Lobron, K. and Nathanson, N. (2005) Assessment of the Bulgarian Enterprise Growth and Investment Project (EGIP), Final Report, USAID/Bulgaria, 2005.
- 37 Poole, A., 1998, Opportunities for change. Structural Survey, 16(4), pp.200-204.
- 38 Birkinshaw, J. and Hood, N., 2000, Characteristics of Foreign Subsidiaries in Industry Clusters. Journal of International Business Studies, 31(1), pp.141-155.

Wealth Effects of Mergers and Acquisitions in the Greek Banking Sector

Aikaterini Markidou¹, Eftychia Nikolaidou², Simone Mateut³

¹SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, amarkidou@seerc.info

²City College, Dept. of Business Administration and Economics, 2 Kalapothaki Str, 54624 Thessaloniki, Greece, enikolaidou@city.academic.gr

³The University of Sheffield, Dept. of Economics, 9 Mappin Str, Sheffield, S1 4 DT, UK, s.mateut@sheffield.ac.uk

This paper provides empirical evidence on the wealth effects of mergers and acquisitions in the Greek banking sector. The Greek banking system has not attracted research interest due to data deficiencies. By employing event study methodology on a sample of 8 bank deals observed over the period 1996-1999, this paper finds that on an aggregate basis M&As of Greek banks create limited value. Specifically, the results indicate that bidding bank shareholders experience insignificant abnormal returns, while abnormal returns are positively related to target bank shareholders offering +7,44% gain over a 12-day [-10;+1] event window. Combined firms earn a modest gain at +2,91%. According to the results, bank mergers in Greece, neither create nor harm shareholder wealth, consistent with prior research performed in the US and Europe.

Keywords

Banks, event study, Greece, mergers and acquisitions.

1. Introduction

Over the past 20 years, increased competition and regulatory changes have fostered a wave of consolidations through mergers and acquisitions in the US, the EU and the global banking industry. In Greece, banks felt the need to follow the international trend of forming larger institutions in the mid '90s, with a long delay compared to the rest of Europe. This consolidation trend raises important questions regarding the immediate and longer-term effects of bank mergers on bank performance. There are numerous studies that investigate the consequences of M&As, however, the vast majority of the existing studies on the issue focus on the financial sector of more developed economies. Given the lack of studies on the Greek banking industry, this study aims to add to the ongoing literature by providing empirical evidence on the value implications of bank M&As in Greece by focusing on eight cases of merged banks.

It is well known in the relative literature that corporate mergers have many motives such as deregulation, globalization, economies of scale and scope, cost reduction, geographic expansion, administrative synergies, financial synergies, tax advantages, diversification of risks. Specifically for bank mergers and acquisitions, there are many factors that contribute and affect them. M&A activity results in overall benefits to shareholders when the consolidated post-merger firm is more valuable than the simple sum of the two separate pre-merger firms. Creation of value can be achieved in several ways: (1) increase of market power, (2) risk diversification, (3) performance improvement following the merger, (4) the cutting of costs through back-office consolidation of systems and increased leverage with technology providers, (5) the achievement of economies of scale and scope, (6) relaxation of geographic constraints and harmonization of regulatory and supervisory environments, (7) replacement of less efficient with more effective managers.

Empirical evidence on the value of bank M&As presents a clear paradox. The results obtained by the majority of researchers are inconclusive of the value effects of bank M&As. Both accounting and event studies in the U.S. and Europe offer no significant evidence of value gains. Yet, mergers continue. Despite the vast amount of literature on the efficiency gains that arise from M&As in the banking sector, there is lack of studies for smaller economies where the trend of M&As is a more recent phenomenon.

This study takes the opportunity to readdress the effects of M&As focusing on the Greek banking industry and offers an important contribution to the ongoing literature by providing evidence from a country that has not attracted any research interest due to data deficiencies. Whether or not Greek bank M&As actually achieve the expected performance gains is the critical question of this paper. Specifically, the paper investigates whether there is any shareholder wealth resulting from bank mergers and acquisitions in the Greek banking industry.

This paper looks at Greek banks involved in mergers and acquisitions over the period 1996-1999. The desire is to explore the stock market reaction to merger announcements based on market data from a sample of 8 bank deals involving 13 banks listed on the Athens Stock Exchange (ASE). To assess the value implications of M&As, this study follows the only approach that directly allows this kind of an assessment: the event study methodology (see [1]). The effects on shareholders of the targets and of the bidders are analyzed both separately and combined. The combined analysis allows to judge on the entire transaction as a whole. The assessment of banking M&As as a whole is important and meaningful because the net wealth creation can be judged upon.

The rest of the paper is structured as follows: Section 2 reviews the literature on bank M&As; Section 3 establishes data requirements and outlines the methodology; Section 4 presents the results while Section 5 summarizes the findings.

2. Literature Review

To date, most of the available knowledge on M&As in banking comes from the US market. On the other hand, the trend of bank M&As has not been adequately examined in Europe. A major drawback is the methodological difficulties in studying the fragmented European markets.

Research literature on the effects of consolidation can be classified: dynamic efficiency studies, operating performance studies and event studies. Dynamic efficiency studies incorporate more sophisticated efficiency techniques in order to estimate the effects of bank consolidation. The most common efficiency estimation techniques are data envelopment analysis (DEA), free disposal hull analysis (FDH), the stochastic frontier approach (SFA), the thick frontier approach (TFA), and the distribution-free approach (DFA). The first two of these are nonparametric techniques and the latter three are parametric methods. Many researchers have examined bank efficiency in terms of scale and scope economies [2],[3],[4]. In bank M&As, increased efficiency of operation is routinely projected and seldom realized. The effects of bank M&As on profit efficiency consequences of consolidation have been investigated by many authors (e.g. [5], [6]). The empirical evidence suggests that the potential for increasing revenue efficiency has been realized, as merging banks have reduced risk through portfolio diversification, enabling them to increase loan activity and enhance profits. Some other studies focus on cost X-efficiency such as those of Berger and Humphrey [7]; Peristiani, [4]; Rhoades [8]; Shaffer [9]; DeYoung [10]. The empirical evidence appears to confirm the potential for increases in cost efficiency, but it also finds that this potential has largely gone unrealized. In Europe, research on efficiency gains incorporates the studies of Vander Vennet [11]; Altunbas et al. [12]; Resti, [13]; and Huizinga et al. [2]. A number of dynamic studies examined the effects of bank consolidation on service availability and small business lending [14],[15],[16]. Findings gauge that larger bank mergers are in general associated with a rapid increase in the level of lending to large businesses resulting in a slow increase or decline in small business lending.

Operating performance studies compare simple pre- and post-merger profitability ratios, such as ROA, ROE based on accounting values. There is no consensus as to whether mergers increase profitability, some of these studies found improved profitability ratios associated with bank mergers [17], [18]

although most others found no improvement in these ratios [19], [20], [21]. Despite the substantial diversity among operating performance studies, the findings point strongly to a lack of improvement in efficiency or profitability as a result of bank mergers.

Study	Regional focus	Years studied	N ^a	Event window in days	CAR Bidder ^b	CAR Target	CAR Comb. Entity ^c
Lobue (1984)	USA	unknown	37	[-24M;+18M]	+24%	N/A	N/A
Bertin et al. (1989)	USA	82-87	33	[-5;+5]	+1,33%	N/A	N/A
Cornett, De (1991)	USA	82-86	152	[-15;+15]	n.s.	+9,66%	N/A
Cornett, Tehranian (1992)	USA	82-87	30	[-1;0]	-0,8%	+8%	N/A
Houston, Ryngaert (1994)	USA	85-91	153	[-4L;+1A] ^d	-2,3%	+14,8%	+0,5%
Madura, Wiant (1994)	USA	83-87	152	[0;36M]	-27,06%	N/A	N/A
Zhang (1995)	USA	80-90	107	[-5;+5]	n.s.	+6,96%	+7,33%
Siems (1996)	USA	1995	19	[-1;+1]	-1,96%	+13,04%	N/A
Pillof (1996)	USA	82-91	48	[-10;0]	N/A	N/A	+1,44%
Subrahmanyam et al.(1997)	USA	82-87	263	[-1 ;+1]	-0,9%	N/A	N/A
Kwan, Einsenbeis (1999)	USA	89-96	3844	[-1;0]	N/A	N/A	+0,77%
Tourani-Rad et al. (1999)	Europe	89-95	17 ;56	[-40 ;+40]	n.s.	+5,71%	N/A
Cybo-Ottone et al. (2000)	Europe	87-98	46	[-10 ;0]	n.s.	+16,1%	+4,0%
Becher (2000)	USA	80-97	558	[-30 ;+5]	-0,1%	+22,6%	+3,0%
Brewer et al. (2000)	USA	90-98	327	[0 ;+1]	N/A	+8,3-14% ^g	N/A
Kane (2000)	USA	91-98	110	{0}	-1,5%	+11,4%	N/A
Hatzigayos et al. (2000)	Greece	98-99	4	[-10;+20]	n.s.	N/A	N/A
Houston et al. (2001)	USA	85-96	64	[-4L ;+1A] ^d	-2,61%	24,60%	+3,11%
Beitel, Schierech (2001)	Europe	85-97	98	[-20;+20]	-0,20%	+16,0%	+1,29%
DeLong (2001)	USA	88-95	280	[-10;+1]	-1,70%	+16,61%	n.s.
Hart, Apilado (2002)	USA	94-97	22	{0}	-0,63%	+5,81%	+0,48%

Table 1 Cumulated abnormal returns for event studies of the last 20 years

^a Number of M&As studied

^b CAR = Cumulated abnormal return; n.s. = not significant; N/A stands for not researched in the study

^c Combined entity of the target and the bidder

^d -4 days prior to the leakage date to 1 day after the announcement

^e No tests for significance

^g The authors study only different sub-samples without presenting results for the entire sample

Source: Update based on Beitel and Schiereck (2000), p. 4.

The basic idea of bank consolidation event studies is to determine if there are any abnormal returns in share prices of the bidders and/or of the targets, and/or of the combined entities around the announcement of a merger or acquisition. In general, findings are not consistent across event studies. For a better understanding of the literature of event studies see Table 1.

The bulk of empirical research shows no evidence of value gains from bank mergers or from increased bank size per se beyond a small size. Cornett and Tehranian [17]; Houston and Ryngaert [22],[23]; Madura and Wiant [24]; Siems [25], Subrahmanyam et al. [26]; DeLong [27]; Becher [28]; Kane [29], Hatzigayos et al.[30], Beitel and Schiereck [31]; and Hart and Apilado [32] studied abnormal returns of acquirers and they found that average cumulative abnormal returns of acquirers were negative around the merger announcement date.

The studies of Cornett and De [33], Zhang [34], Tourani-Rad and Van Beek [35], Cybo-Ottone and Murgia [36] present no significant value creation in the bidder share prices. Also of note is, that there are some event studies conducted in the 1980s reporting positive effects in the stock price of acquiring banks. For example, Lobue [37] finds that acquiring banks capture significant positive abnormal returns suggesting that particular properties of the target firm, such as New York Stock Exchange listing, relative size of sellers and of bidders, and state branching laws are key explanations for positive bidder returns. Bertin et al. [38] focusing on failed bank acquisitions show that bidding banks actually underbid the value of the failed targets.

It is worth noting that all event studies looking at the market's reaction to proposed mergers regarding target banks show the same results. Acquired firm shareholders gain at the expense of the acquiring firm. This is documented over the course of many studies covering different time periods, different locations, and different size sample.

Analysis of merger gains examining stock price performance of the bidder and target firm around the announcement of a merger or acquisition indicate that overall wealth effects from bank mergers are positive over time [20], [21], [28], [23], [32].

Although European research on bank efficiency has not matched the volume of US studies this has began to change in recent years. There is some evidence that M&As in Europe increase combined value. A notable study of the European market is the recent work by Cybo-Ottone and Murgia [36], who documented that there is a positive and significant increase in stock market value for the targets and the combined entity at the time of the deal announcement. It should be noted that the sample used also contained cross-product deals in which banks expand into insurance or investment banking, since regulations allow EU banks to offer both banking and insurance products. Beitel and Schiereck [31] also studied value creation of European banking consolidation and reported positive findings for the combined entity and for the shareholders of the targets that earn considerable and significant positive abnormal returns. The results for the shareholders of the bidders are insignificantly negative. Tourani-Rad and Van Beek [35] found that shareholders of the targets experience significantly positive returns while abnormal returns for the bidding banks are very modest and not statistically significant due to the relative small size of the target comparing to that of the bidder.

As far as M&As in the Greek banking sector as concerned, to our knowledge, the paper of Hatzigayos et al. [30] is the first study that examines the consolidation of listed banks in the Greek market. The authors investigate 4 bank deals over the period 1998-99 when the first merger wave took place in Greece. The results point at insignificant negative abnormal returns for the bidding banks at a merger announcement mainly due to overpriced takeovers.

In summary, the event studies generally find that stockholders of target firms have gains. However, the evidence regarding returns to bidders has mixed results, thus, is too inconsistent to permit any clear conclusion. Net wealth analysis around the announcement of the deal results in overall small but statistically significant gain from mergers, combining the positive abnormal returns on the shares of the target bank and the negative abnormal returns on the shares of the bidder. On balance, then, evidence from the studies undercut the hypothesis that the financial markets expect mergers to improve bank performance. It becomes clear, then, that the extensive empirical literature on bank M&As, the different methodologies employed along with the developments in the various statistical and econometric techniques have not succeeded in providing a clear-cut answer on the effects of bank

M&As. Evidence is still mixed and further studies are needed, in particular for less advanced/developed countries, given the gap in the literature.

3. Data sample and methodology

3.1. Data sources and data sample

Prior to discussing data sources and the various data problems encountered, it is necessary to provide a brief overview of the Greek banking sector. Commercial banks have been the dominant group in the Greek banking system. In 1995, besides the Central Bank, there were 31 credit institutions established and operating in the Greek banking market, comprising 20 commercial banks and 11 specialized credit institutions (2 investment banks, 3 housing banks, 1 savings bank, 1 specific purpose bank, and 4 credit co-operatives). A specific feature of the Greek commercial banking system is the dominant role of a few large banks, most of them state-owned, having considerable market power [39]. The Greek banking situation differs from the rest European situation in terms of number of banks operating domestically. European banking systems are characterized by fragmented local markets, where small credit cooperative banks usually covering a very small area, with just one or few branches. In 1999, the total number of commercial banks operating in major European countries was: 362 commercial banks in France, 203 in Germany, 146 in Spain, 237 in Italy and 41 in UK [40].

Financial liberalization has allowed the entrance of new, mainly small, private banks and led to an increase in the number of branches, resulting in a decline in the, albeit high, concentration ratio of the commercial banking sector.

The first merger that took place in Greece was that of National Mortgage and National Housing, in the end of 1996. The government was controlling directly the negotiations, since there was no legal framework for bank M&As adjusted for hostile takeovers. The legal framework was finally expanded by the end of 1999 by a Presidential Decree. As a result, the phenomenon of merger wave started in the Greek banking market [30]. After all the mergers and acquisitions in the banking industry today the number of commercial banks operating in Greece has decreased considerably.

As mentioned earlier, M&As in the Greek banking industry have started only after 1996 as a result of state intervention, which for a long time hindered competition and created a distorted market environment. Thus, the number of M&As included in the sample is very limited compared to that of US and European studies. To identify M&A transactions of Greek banks between 1996-1999, this study relies on two data sources: The Stock Exchange Center of Thessaloniki and the Economic Bulletins of Commercial Bank. The Stock Exchange Center of Thessaloniki provides individual equity values (historical data for stock prices of banks involved in M&As), banking industry and market returns. The exact announcement dates of M&As are not readily available (the Economic Bulletins of Commercial Bank provides only yearly tables of M&As in Greek banks), thus a lot of research on financial newspapers was needed. For the analysis of additional data (e.g. total assets, total equity) based on bank balance sheets and income statements, the study relies on financial statements of the Greek banking system provided by the Hellenic Bank Association (HBA) that cover data only for the period 1995-1999. Overall, the collection and tidying of the required data proved to be a time-consuming procedure, since in Greece there is no specific institution dealing with bank M&As like Computasoft M&A data or Thompson Financial SCD (Securities Data Company-Mergers and Acquisitions Database) in Europe. Despite the lack of readily available data, we managed to gather all relevant information and to construct the data sample required for the empirical analysis.

Specifically, the sample consists of banks that satisfy the following criteria:

- Both, the bidding and the target banks are publicly traded banking institutions listed on the Athens Stock Exchange (ASE) for at least 252 trading days (a full year) prior to the announcement and 20 days after the announcement of a merger transaction.
- The merger or acquisition must have occurred before 1999.

- Both of the merged banks must be healthy institutions at the time of the merger.
- The transaction has been closed – the deal status hence is “completed”.
- The M&A deal is a full merger of the two banks or entails the transfer of control from the target to the acquiring bank.

A total number of fifteen bank mergers and acquisitions took place between 1996-1999. Seven mergers were eliminated from the sample as they did not satisfy the above criteria (ie. in four cases, the bidding or target banks were not publicly traded banking institutions, which means that there were no share prices to perform event study methodology and in three cases, Greek banks involved in the take-over of network of foreign banks).

Total Assets (million drs) at Start of Merger Year								
Year	Number of mergers	Acquirer			Target			Relative Size (%) ^a
		Mean	Minimum	Maximum	Mean	Minimum	Maximum	Mean
1997	1	2.280.559	2.280.559	2.280.559	123.756	123.756	123.756	5,15%
1998	5	2.251.109	128.726	9.802.976	726.245	119.520	2.515.763	40,70%
1999	2	2.782.274	1.290.418	4.274.130	1.804.520	1.729.332	1.879.707	43,91%
Full sample	8	2.387.582	128.726	9.802.976	920.502	119.520	2.515.763	37,06%

Table 2 Summary statistics of mergers and acquisitions in sample

Notes: The sample consists of deals taking place from 1997 to 1999 involving eight publicly traded banking institutions

^a Relative size equals target total assets divided by target plus acquirer total assets, with assets measured at the start of merger year.

Total Equity (million drs) at Start of Merger Year								
Year	Number of mergers	Acquirer			Target			Relative Size (%) ^a
		Mean	Minimum	Maximum	Mean	Minimum	Maximum	Mean
1997	1	82.530	82.530	82.530	7.106	7.106	7.106	7,93%
1998	5	103.475	9.540	380.874	29.864	6.167	93.407	30,58%
1999	2	270.300	189.690	350.909	94.803	60.296	129.310	27,60%
Full sample	8	142.563	9.540	380.874	43.254	6.167	129.310	37,00%

Table 3 Summary statistics of mergers and acquisitions in sample

Notes: The sample consists of deals taking place from 1997 to 1999 involving eight publicly traded banking institutions

^a Relative size equals target total assets divided by target plus acquirer total assets, with assets measured at the start of merger year

Tables 2 and 3 summarize the key features of the sample on a year-by-year basis. As expected, acquiring banks are significantly larger, in terms of total assets, than target banks. The mean asset size of acquirers over the period is 2.387 million drs while the mean target is 920 million drs. On the other hand, the mean equity size for the bidders is 142 million drs while the mean equity size for the targets is 43 million drs. These sizes are substantially larger than the mean acquirer size and mean target size of many US studies (See Table 4). More impressive is the fact that the relative size of the targets is considerably larger than any other study concerning European or USA bank M&As as it can be seen in Table 4. Even if this paper examines the smallest sample of mergers, just 8 bank deals, the mean

acquiring bank has assets of \$57 billions while the mean target bank has \$22 billion in assets. Cybo-Ottone and Murgia [36] documented that European banks are bigger players than the US peers and this seems to be valid in this study for Greece too.

Statistics	This study (2005)	Selected prior studies					
		BS (2001)	CO (2000)	S (1996)	P (1996)	Z (1995)	HR (1994)
Number of transactions	8	98	46 ^a	19	48	107	153
Study focus	Greece	Europe	Europe	USA	USA	USA	USA
Deal value ^b	N/A	3.343	1.612	2.774	N/A	N/A	N/A
Total assets ^b							
Bidders	57	182	106	61	13	14	N/A
Targets	22	37	24	19	4	2	N/A
Targets in % of Bidders	37,06%	20,4%	22,4%	23,5%	26,1%	14,9%	15,9%

Table 4 Comparison to selected studies

Legend: BS= Beitel and Schiereck [31], CO = Cybo-Ottone and Murgia [36], S = Siems [25], P = Pillof [20], Z = Zhang [34] and HR = Houston and Ryngaert [22].

^a 46 completed transactions. In addition, Cybo-Ottone and Murgia [36] analyzed 8 cancelled deals.

^b In billions of USD.

Source: Update based on Cybo-Ottone and Murgia [36] p. 838.

3.2. Methodology

To see whether the market anticipates and prices any of the acquisitions based on initial characteristics of the targets or acquirers or in other words to examine whether changes in performance are anticipated and priced, a standard event study (see [1]) is performed, where differences in the stock returns between acquiring banks or target banks and the market are used as estimates of abnormal or excess returns for a 12-day window [-10;+1] around the merger announcement date, using the following model:

$$AR_{it} = R_{it} - (a_i + b_i R_{mt}) \quad (1)$$

The parameters of equation (1), a_i and b_i , are estimated from a market model, as follows:

$$R_{it} = a_i + b_i R_{mt} + e_{it} \quad (2)$$

Chart 1 displays a time line that shows the market model estimation period, the event date, and the periods used to compute abnormal returns.

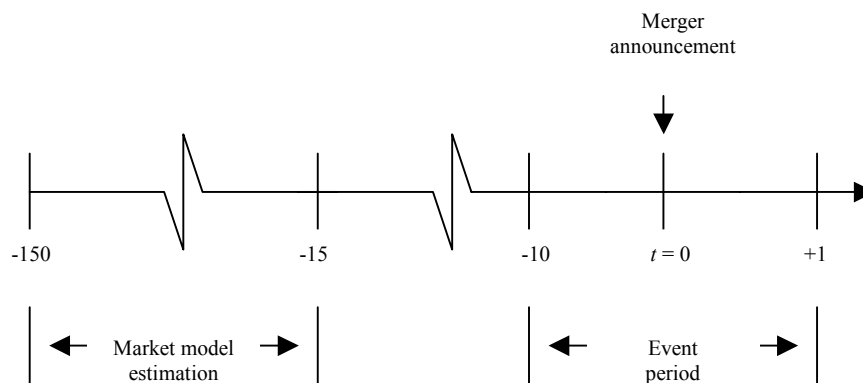


Chart 1 Event study over time line

Most studies examine the abnormal returns of acquirers and targets separately (See Table 1), but several papers analyze the total change in shareholder wealth. In such cases, the value-weighted sum of acquirer and target abnormal returns is the appropriate measure of overall gains stemming from merger and acquisition activity. This measure quantifies the value reaction that the market believes the merger will provide because false interpretations can be made when looking solely at the outcomes of the bidder or the target. Cumulative abnormal returns of the combined entity (bidder and target firms together) are calculated by following the method outline in Houston and Ryngaert [22]:

$$\text{Combined Cumulative Abnormal Returns} = \frac{(CAR_{ib} V_{ib}) + (V_{it} CAR_{it})}{(V_{ib} + V_{it})} \quad (5)$$

where V_i is the value of the bank's stock -10 days before the merger announcement date for the bidder and target respectively over the 12-day window. To gauge statistical significance, a z-test and subsequent p-value are calculated from the mean assuming a normal distribution using the suggestions described in Dodd and Warner [41].

4. Empirical results

Following the methodology outlined in the previous section, several event windows are used to calculate abnormal returns ranging in size from twelve days, spanning days $[t = -10, t = +1]$ to only two days $[t = 0, +1]$. Table 5 provides the cumulative abnormal returns for bidders.

<i>Bidders</i> (N = 8) Event window	CAR in % ^a	Pos.	Neg.	Z-test	p-value
[-10;0]	1,74	4	4	0,01	0,25477
[-5;0]	1,88	3	5	0,03	0,19548
[-2;0]	2,54	5	3	0,04	0,20358
[-1;0]	0,08	4	4	0,25	0,22571
{0}	-0,78	2	6	0,50	0,11929
[-1;+1]	-0,88	4	4	0,22	0,11271
[0;+1]	-1,74	4	4	0,47	0,29943
[-10;+1]	0,78	3	5	0,31	0,33732

Table 5 Cumulative abnormal returns (CARs) of the acquiring banks in Greece between 1996-1999.

Notes: This table presents the results for an event study examining 8 targets from Greek bank M&As. Abnormal returns were calculated using OLS-regression. OLS parameters have been estimated for a period of 135 trading days prior to the event window $[-10;+1]$. As market returns we applied ASE indice (Athen Stock Exchange). Tests of significance are calculated from standardized abnormal returns employing the Dodd-Warner [41] procedure.

^a ***=significant at the 1 percent level, **=significant at the 5 percent level, *=significant at 10 percent level.

In general, prior to the merger announcement date, bidders experience positive returns. Over the 11-day window $[-10;0]$, bidder CARs are accounted for +1,74%, while the 3-day window $[-2;0]$ offers +2,54% gains for the shareholders of the acquiring firms. However, this trend seems to be altered exactly on the announcement date where bidder abnormal returns fall significantly. This is very clear in the 2-day event window $[0;+1]$ where the losses for bidders reach $-1,74\%$. Overall, we find a positive and statistically insignificant abnormal returns to acquiring firms amounting to a twelve-day cumulative abnormal return of only +0,78%, a very modest average gain. One explanation for this slight increase in returns for acquiring banks is the fact that the considerable size of target banks in Greece along with their strong financial performance do not allow bidding firms to exploit any significant gains from efficiency increase and cost savings. However, the results are not seriously differentiated with these of an earlier event study by Hatzigayos et al [30]. Their findings indicate that there is an insignificant negative reaction for shareholders of the acquiring firms around the

announcement of a bank merger in Greece. The authors find a non-significant negative reaction of $-0,3\%$ on days -1 to $+5$ after the announcement date. Nevertheless, the sample used in their work is somewhat smaller than that used in this study and the authors computed abnormal returns only for the bidders.

Other previous European studies that look at the returns to bidders report insignificant findings for the shareholders of the acquiring firms. The results of Cybo-Ottone and Murgia [36], Beitel and Shierech [31] and Tourani-Rad and Van Beek [35] are basically the same. However, studies focusing on US M&As indicate significant negative cumulative abnormal returns. Hart and Apilado [32] show $-0,63\%$ losses for a one-day event window $\{0\}$. In addition, DeLong [27] finds $-1,70\%$ return for a twelve-day window $[-10;+1]$, while Houston et al. [23] report $-2,61\%$ return for acquiring firms. The findings for the bidders in this study seem to contradict the findings of major US studies, while tend to confirm several studies conducted in European banking markets indicating neither success nor failure of value creation for the shareholders of acquiring banks.

Targets (N = 8) Event window	CAR in % ^a	Pos.	Neg.	Z-test	p-value
$[-10;0]$	5,43***	5	3	0,96	0,00000
$[-5;0]$	3,76***	4	4	0,86	0,00000
$[-2;0]$	4,54***	6	2	0,29	0,00000
$[-1;0]$	2,72***	4	4	0,39	0,00000
$\{0\}$	1,14***	3	5	0,50	0,00000
$[-1;+1]$	4,73***	3	5	0,72	0,00000
$[0;+1]$	3,15***	4	4	0,67	0,00000
$[-10;+1]$	7,44***	5	3	0,58	0,00000

Table 6 Cumulative abnormal returns (CARs) of targeted banks in Greece between 1996-1999.

Notes: This table presents the results for an event study examining 8 bidders from Greek bank M&As. Abnormal returns were calculated using OLS-regression. OLS parameters have been estimated for a period of 135 trading days prior to the event window $[-10;+1]$. As market returns we applied ASE indice (Athen Stock Exchange). Tests of significance are calculated from standardized abnormal returns employing the Dodd-Warner [41] procedure.

^a ***=significant at the 1 percent level, **=significant at the 5 percent level, *=significant at 10 percent level.

Cumulative abnormal returns for targets across event windows are reported in Table 6. There is no much to say about target returns. Like previous European and US studies, target banks in Greece have positive wealth effects in all event windows. As can be noted observing p-value of the z-test, all measures of CARs are highly significant. This work finds a statistically significant cumulative return $+7,44\%$ for the event window $[-10;+1]$. According to Beitel and Schiereck [31], in Europe, cumulative abnormal returns for targets account for $+16,0\%$ in a 41-day window $[-20;+20]$. Cybo-Ottone and Murgia [36] also find $+16,1\%$ gains for target banks, while Tourani-Rad and Van Beek [35] show $+5,71\%$ wealth increase in a 81-day event window $[-40;+40]$. The same results are found in all studies performed in USA too. Targets experience superior performance regardless of the days studied in the event windows. In other words, M&As in Europe and USA act in favor of target's shareholders. This outcome suggests that target management and shareholders may prefer to withdraw from deals where there are no significant opportunities to exploit merger gains.

The results of the event study for the combined entity are given in Table 7. Examining simultaneously both the acquiring and targeted banks, allows us to determine whether bank M&As create rather than transfer wealth. The market reaction for the combined entity to a merger announcement for several days surrounding the merger announcement shows a slight increase in the combined abnormal returns for 8 pairs of acquiring and targeted banks in sample. Table 7 indicates that over the 11-day window $[-10;+1]$, cumulative abnormal returns to the combined entity are $+2,91\%$. Positive returns to targets are essentially offset by insignificant returns to bidders. It is interesting to note, however, that this result is consistent with accounting-based studies that provide evidence for limited efficiency gains from bank mergers [32], [21], [20]. Comparing the results of this study with those reported in Table 1, Cybo-

Ottone and Murgia [36] find +4,0% increase in the market value for the combined entity in a sample of 46 European bank mergers.

Combined entity (N = 8) Event window	CAR in % ^a	Pos.	Neg.	Z-test	p-value
[-10;0]	1,10***	5	3	0,30	0,00056
[-5;0]	0,24***	4	4	0,22	0,00099
[-2;0]	1,08***	6	2	0,04	0,00268
[-1;0]	0,85***	4	4	0,23	0,00044
{0}	0,44***	3	5	0,50	0,00003
[-1;+1]	2,42***	3	5	0,42	0,00011
[0;+1]	1,15***	6	2	0,58	0,00005
[-10;+1]	2,91***	3	5	0,29	0,00413

Table 7 Combined cumulative abnormal returns (CARs) from bank takeovers in Greece between 1996-1999.

Notes: This table presents the results for an event study examining 8 targets from Greek bank M&As. Abnormal returns were calculated using OLS-regression. OLS parameters have been estimated for a period of 135 trading days prior to the event window [-10;+1]. As market returns we applied ASE indice (Athen Stock Exchange). Tests of significance are calculated from standardized abnormal returns employing the Dodd-Warner (1983) procedure.

^a ***=significant at the 1 percent level, **=significant at the 5 percent level, *=significant at 10 percent level.

Beitel and Schiereck [31] also studied mergers in Europe, show +1,29% increase in combined value. Studies on the wealth effects of US bank M&As such as those of Houston et al. [23]; Becher [28]; Zhang [34]; Houston and Ryngaert [23] find that mergers can create little value on a net and aggregate basis. According to the aforementioned studies, this work is consistent with actual measured performance gains and the bulk of European and US event studies. Summing up, given these results, it is unclear if the market expects mergers to lead to gains in performance when the mean overall change in shareholder value is quite small. For a more complete picture of the CARs during the investigation period for the bidders, the targets as well as for the combined entity, see Figure 1.

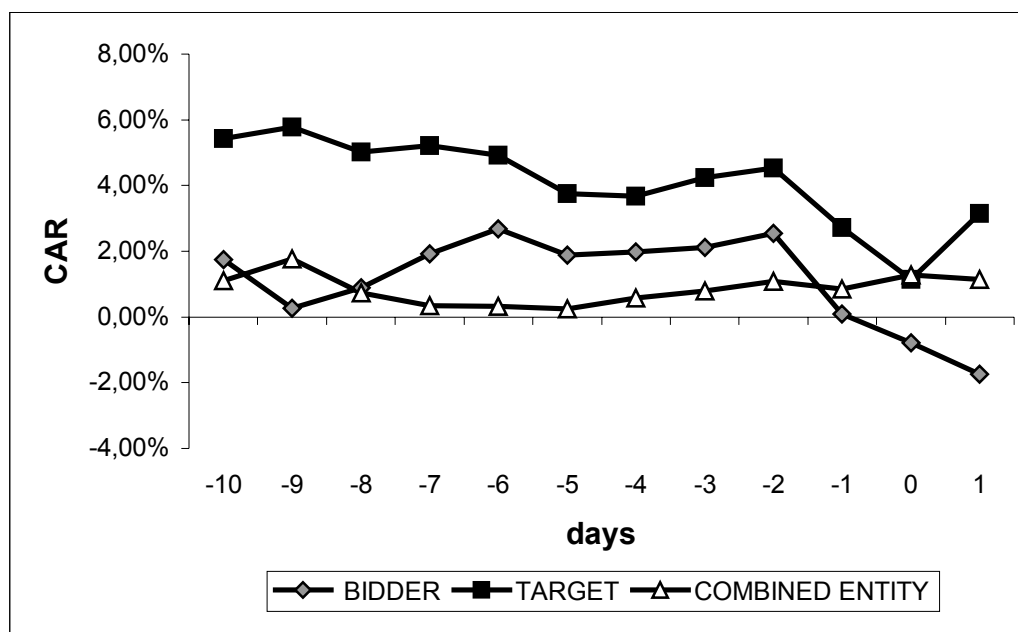


Figure 1 CARs for the whole banks sample

5. Conclusion

This study has undertaken a comprehensive empirical analysis of the wealth effects of bank M&As in Greece over the period 1996-1999. The focus of the analysis has been Greece given the very limited amount of research on the particular topic for this country. This study employed the event study methodology and provided empirical evidence for 8 cases of Greek bank M&As. Results suggest that there are insignificant abnormal returns for acquirers, significant positive abnormal returns at +7,44% for targets and +2,91% for the combined entity, in the event window [-10;+1]. The findings assume that, on average, the Greek bank mergers neither create nor destroy shareholder wealth. The outcome of this work is consistent with the findings of the only Greek event study of Hatzigayos et al. [30] and the majority of European and US literature on M&A wealth effects.

The literature on the value of bank mergers and acquisitions presents a clear paradox. Empirical evidence cannot indicate clearly that, on average, there is statistically significant gain in value or performance from merger activity. On average, acquired firm shareholders gain at the expense of the acquiring firm and market value of the combined entity appears to have little improvement around the announcement of the transaction. This is documented over the course of many studies covering different time periods and different locations. Yet, mergers continue. Indeed the merger wave that has swept the US in the '80s and early '90s has been transferred particularly in West Europe and recently in less developed countries like Greece. Since the merger activity trend has recently begun in Greece, when more data become available, future research should employ dynamic analysis methods by focusing in cross-sectional econometric analysis to determine which factors affect wealth creation through mergers.

References

- 1 Brown, S.J., Warner, J.B. Using daily stock returns – The case of event studies. *Journal of Financial Economics* 1985; 14: 3-31.
- 2 Huizinga, H.P., Nelissen, J.H.M., and Vander Venet, R. Efficiency effects of bank mergers and acquisitions in Europe. Tinbergen Institute, Amsterdam 2001; DP-01-88.
- 3 Hancock, D., Humphrey, D.B., Wilcox, J.A. Cost reductions in electronic payments: The roles of consolidation, economies of scale, and technical change. *Journal of Banking & Finance* 1999; 23: 391-421.
- 4 Peristiani, S. Do mergers improve the cross-efficiency and scale of U.S. banks? Evidence from the 1980s. *Journal of Money, Credit, and Banking* 1997; 29: 326-337.
- 5 Benston, G.J., Hunter, W.C., Wall, L.D. Motivations for bank mergers and acquisitions: Enhancing the deposit insurance put option versus earnings diversification. *Journal of Money, Credit and Banking* 1995; 27: 777-788.
- 6 Akhavein, J.D., Berger, A.N., Humphrey, D.B. The effects of megamergers on efficiency and prices: Evidence from a bank profit function. *Review Industrial Organization* 1997; 12: 95-139.
- 7 Berger, A.N., Humphrey, D.B. Bank scale economies, mergers, concentration, and efficiency: The U.S. experience. Wharton Financial Institutions Center, University of Pennsylvania 1994; WP-94-25.
- 8 Rhoades, S.A. Efficiency effects of horizontal (in-market) bank mergers. *Journal of Banking & Finance* 1993; 17: 411-422.
- 9 Shaffer, S. Can megamergers improve bank efficiency?. *Journal of Banking & Finance* 1993; 17: 423-436.
- 10 DeYoung, R. Bank mergers, X-efficiency, and the market for corporate control. *Managerial Finance* 1997; 23: 32-47.
- 11 Vander Venet, R. The effects of mergers and acquisitions on the efficiency and profitability of EC credit institutions". *Journal of Banking & Finance* 1996; 20: 531-558.
- 12 Altunbas, Y., Molyneux, P., Thornton, J. Big bank mergers in Europe: An analysis of the cost implications. *Economica* 1997; 64: 317-329.
- 13 Resti, A. Regulation can foster mergers, can mergers foster efficiency? The Italian case. *Journal of Economics and Business* 1998; 50: 157-169.

- 14 Strahan, P.E., Weston, J.P. Small business lending and the changing structure of the banking industry. *Journal of Banking & Finance* 1998; 22: 821-845.
- 15 Peek, J., Rosengren, E.S. Bank consolidation and small business lending: It's not just bank size that matters. *Journal of Banking & Finance* 1998; 22: 799-819.
- 16 Berger, A.N., Saunders, A., Scalise, J.M., Udell, G.F. The effects of bank mergers and acquisitions on small business lending. *Journal of Financial Economics* 1998; 50: 187-229.
- 17 Cornett, M.M., Tehranian, H. Changes in corporate performance associated with bank acquisitions. *Journal of Financial Economics* 1992; 31: 211-234.
- 18 Spindt, P.A., Tarhan, V. Performance economies associated with small and medium sized bank mergers. *Journal of Banking & Finance* 1993; 17: 460-462.
- 19 Linder, J.C., Crane, D.W. Bank mergers: Integration and profitability. *Journal of Financial Services Research* 1992; 7: 35-55.
- 20 Pillof, S.J. Performance changes and shareholder wealth creation associated with mergers of publicly traded banking institutions. *Journal of Money, Credit and Banking* 1996; 28: 294-310.
- 21 Kwan, S., Eisenbeis, R.A. Mergers of publicly traded banking organizations revisited". *Economic Review*, Federal Reserve Bank of Atlanta 1999; 84: 26-37.
- 22 Houston, J.F., Ryngaert, M.D. The overall gains from large bank mergers. *Journal of Banking & Finance* 1994; 18: 1155-1176.
- 23 Houston, J.F., James, C.M., Ryngaet, M.D. Where do merger gains come from? Bank mergers from the perspective of insiders and outsiders. *Journal of Financial Economics* 2001; 60: 285-331.
- 24 Madura, J., Wiant, K.J. Long-term valuation effects of bank acquisitions. *Journal of Banking & Finance* 1994; 18: 1135-1154.
- 25 Siems, T.F. Bank mergers and shareholder wealth: Evidence from 1995's megamergers deals. Federal Reserve Bank of Dallas, *Financial Industry Studies* 1996, August.
- 26 Subrahmanyam, V., Ranga, N., Rosenstein, S. The role of outside directors in bank acquisitions. *Financial Management* 1997; 26: 23-36.
- 27 DeLong, G. Stockholder gains from focusing versus diversifying bank mergers. *Journal of Financial Economics* 2001; 59: 221-252.
- 28 Becher, D.A. The valuation effects of bank mergers. *Journal of Corporate Finance* 2000; 6: 189-214.
- 29 Kane, E.J Incentives for banking megabank mergers: What motives might regulators infer from event-study evidence. *Journal of Money, Credit and Banking* 2000; 32: 671-701.
- 30 Hatzigayos, T., Lyroudi, K., Subeniotis, D. The legal innovations and the practical experience of mergers in the Greek banking system. *Financial stabilization and economic growth forum*, 26-27 October 2000, Svishtov, Bulgaria.
- 31 Beitel, P., Schiereck, D. Value creation at the ongoing consolidation of the European banking market. *Institutions for mergers and acquisitions*, Germany 2001, WP-05-01.
- 32 Hart, J.R., Apilado, V.P. Inexperienced banks and interstate mergers. *Journal of Economics and Business* 2002; 54: 313-330.
- 33 Cornett, M.M., De, S. Common stock returns in corporate takeover bids: Evidence from interstate bank mergers. *Journal of Banking & Finance* 1992; 15: 273-295.
- 34 Zhang, H. Wealth effects of US bank takeovers. *Applied Financial Economics* 1995; 5: 329-336.
- 35 Tourani Rad, A., Van Veeck, L. Market valuation of European bank mergers. *European Management Journal* 1999; 5: 532-539.
- 36 Cybo-Ottone, A., Murgia, M. Mergers and shareholder wealth in European banking. *Journal of Banking & Finance* 2000; 24: 831-589.
- 37 Lobue, M. Categorical bank acquisitions. *Journal of Bank Research* 1984; Winter: 274-282.
- 38 Bertin, W.J., Ghazanfari, F., Torabzadeh, K.M. Failed bank acquisitions and successful bidders' returns. *Financial Management* 1989; Summer: 93-99.
- 39 Hondroyiannis, G., Lolos, S., Papapetrou E. Assessing competitive conditions in the Greek banking system. *Journal of International Financial Markets, Institutions and Money* 1999; 9: 377-391.
- 40 OECD, *Bank Profitability* 1999.
- 41 Dodd, W., Warner, J.B. On corporate governance: A study of proxy contests. *Journal of Financial Economics* 1983; 11: 401-438.

Development of Clusters in South-East Europe: the Case of Bosnia and Herzegovina - a Literature Review

Amira Vejzagic-Ramhorst¹, Panayiotis H. Ketikidis^{1,2}, S.C. Robert Huggins³

¹SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, {aramhorst, pktikidis}@seerc.info

²City College, Dept. of Computer Science, 13 Tsimiski Str, 54624 Thessaloniki, Greece, ketikidis@city.academic.gr

³The University of Sheffield, Management School, 9 Mappin Street, Sheffield, S1 4DT, UK, rhuggins@sheffield.ac.uk

Policies and programs that facilitate clusters have gained increasing importance in the discussion about economic growth in the countries of South-East Europe (SEE) that are undergoing transition from socialist to market economy. Proximity however was not an asset in socialism and policies that promote clusters in these economies face the “challenge” of this socialist legacy as well as additional challenges of transition. This paper aims to give an overview of the existing literature on clusters and various aspects related to clusters such as inter-firm networks and innovation with the intention to set out an agenda of relevant challenges for policies that support cluster development in the region of SEE. Particular focus is given to Bosnia and Herzegovina (BiH).

Keywords

Clusters, innovation, inter-firm networks, social capital.

1. Introduction

Over the last decade clusters have gained increasing importance in the discussion about economic growth. This paper reviews the major contributions in literature on clusters and innovation to set the base for an agenda of issues concerning the cluster policy in SEE economies and in Bosnia and Herzegovina in particular.

2. Context/ Background of the paper

Research on the sources of economic growth has evolved substantially in the last fifty years since the Solow model that stressed physical capital as source of growth. In contrast, the endogenous growth theory (as originally developed by Romer) challenges the pivotal role given to physical capital in promoting growth and considers “total factor productivity” (TFP) as the engine of economic growth more generally. As the development theory moved away from exclusive focus on capital, institutions as the key element for fostering growth are also emphasized [1]. Along with emphasis on institutions, various theories have put learning at the centre of economic development. Knowledge creation and assimilation have become central for understanding firm performance and economic growth.

Researchers have argued that there is a strong link between economic growth and clusters. Their argument is that since clustering allows firms to experience higher TFP growth than otherwise, the existence of a cluster allows a region or a country to have faster economic growth than it would otherwise have in absence of clusters. Literature identifies several potential sources of higher

productivity growth for firms operating in clusters [2]. These include the benefits stemming from the localized external economies (such as access to a common pool of resources, in particular specialized labour and a dedicated network of suppliers and customers which minimizes search costs) but also reduction in transaction costs, reduced costs via effective learning, innovation and technological development. Finally, there are also increasing returns to firms driven by systemic properties embedded within the local systems, first mover advantages from the initial territorial specialization and advantages to being customer driven organizations [2]. Propriis and Driffield [3] have also analyzed the importance of clusters in terms of knowledge spillovers from FDI, but also for “reverse spillovers”.

When thinking about ways to encourage economic growth in transitional economies of SEE policy makers and researchers have largely focused on reforms prescribed by the “Washington Consensus”, including removal of barriers to trade and interdependencies between trade, FDI and growth. While there is a body of research on evidence of a positive long-run correlation between structural reform and economic growth [4], the elusive mechanism which translates reforms into technology integration and growth in the SEE region needs to be better understood. Gristock [5] argues therefore that the key question is “to better understand how countries can develop capabilities to assimilate, generate and manage technical change and learning”. Radosevic [6] extends Gristock’s argument of importance of learning and innovation to growth and posits that development of the region as a locus of innovation is essential for the restructuring and sustainable growth in SEE.

Over the last several years clusters have gained increasing importance in the discussion about fostering economic growth in SEE. Experience from developed and developing countries has been translated into a wide range of cluster policies and supporting programs supported by individual national and regional governments and different donor agencies in SEE. Given challenges specific to transitional economies and limited empirical research on clusters in SEE countries, there is a need to develop a greater understanding of the specific context for policies that target clusters in SEE countries.

3. Overview of literature

The concept of specialised industrial localization is hardly new [7]. As early as 1891 Alfred Marshall introduced the concept of “concentrations of specialized industries in particular localities” and elaborated on economic benefits that accrue to firms primarily because of their geographic proximity. These benefits are referred to as Marshallian externalities or Marshall-Arrow-Romer (MAR) externalities and refer to intra-industry economies of localization and are often grouped as [2]:

- Economies of specialization: a localized industry can support a greater number of specialized suppliers, thus obtaining a greater variety at lower cost
- Labour market economies: localized industries attract and create pools of workers with specialized skills
- Knowledge spillovers: information about novelties flows more easily among agents located within the same area thanks to social bonds and frequent face to face contacts.

In the past three decades there has been a strong renewed attention to the writings of Alfred Marshall on industrial localization. Research on the model of industrial organization based on agglomeration of enterprises has expanded in many branches to include economic sociology, industrial economics, regional economics, development economics and international business theory [2]. There is now a vast body of published research on clusters based on various strands of theoretical literature leading to different approaches to studying clusters. Innovation and knowledge diffusion are also an integral element in cluster research. Another related field of research concerns the role of social capital and trust in clusters. In terms of empirical research, there is a substantive body of research on different clusters in USA and Europe. Though not as nearly as substantive, there are also several case studies of clusters in developing economies. More recently there has also been growing research on clusters in transition economies with the focus on clusters in Central Europe [8; 9; 10]. However, as far as the

region of South-East Europe is concerned, depending on the specific country in focus, research on clusters is either non-existent or very scarce.

3.1 Clusters

3.1.1 Definition

Literature review reveals that there is agreement on the lack of agreement on the definition of clusters [2; 3; 7; 11; 12; 13]. Van Dijk and Sverrisson [14] point out “the cluster concept has been defined in multiple ways by researchers and therefore they also tend to come up with different selection of characteristics according to the premise that you find what you are looking for”. Martin and Sunley [7] have argued that the concept of cluster has become a “chaotic” concept too fuzzy to be pinned down. Similarly, Perry [13] notes that “the prominence of clusters is helped by definitional vagueness”.

Table 1, adapted from Martin and Sunley [7] and Belussi [2] highlights the diversity of cluster definitions in cluster literature and also that sometimes the term cluster is synonymously used with the term industrial district.

Rocha and Sternberg [11] argue that the literature on clusters shows that there are three necessary dimensions: geographical proximity, an interfirm network and an inter-organizational or institutional network.

Torre and Rallet [15] distinguish between two types of proximity: geographic and organized proximity and based on intersection of these two types of proximity they provide a grid analysis of different models of geographic organizations. In this framework clusters are characterized by existence of both types of proximity. In clusters, organized proximity, defined by the intensity of client supplier relationships, is based on the co-localization of firms within a zone. Furthermore, Torre and Rallet [15] argue that organized proximity often exists without geographical proximity and that it offers a powerful mechanism of long-distance coordination. Finally, Torre and Rallet [15] also highlight possible negative effects of geographical proximity, arguing that in the case of limited supply goods, geographical proximity can create tensions between firms and can damage their relational fabric. However, the negative externalities of geographic proximity can be limited through processes of cooperation or negotiation.

3.1.2 Typologies

Very much like for the definition of clusters, literature review reveals that there is no single widely accepted cluster typology and that “there are as many typologies as case studies” [2]. It is therefore beyond the scope of this paper to present all typologies from the literature reviewed, rather a few typologies that are frequently referenced will be highlighted.

Gordon and McCann [16] distinguish between three basic forms of clustering: the classic model of pure agglomeration, the industrial-complex model and the social network model. They highlight that the first two models have developed from the (neo) classical tradition of economics while the third one is based more on the sociological approach. The model of pure agglomeration draws on Marshall’s three reasons why firms would localize in a same area, which include development of a local pool of specialized labour, the increased local provision of non-traded input specific to an industry and maximum flow of information. As Gordon and McCann [16] stress the key to this model is that economic benefits accrue to firms primarily because of their geographic proximity and that this model does not presume any form of cooperation between firms beyond what is in their individual interest. In literature this model/phenomena is also referred to as “economies of agglomeration” and “external economies”.

The second model that Gordon and McCann [16] identify, the industrial-complex model, is characterized by sets of identifiable relations, primarily in terms of trading links, among firms. Therefore, Gordon and McCann [16] explain that in this model the only reason why we might observe spatial industrial clustering is that individual firms, seeking to minimize spatial transaction costs, have located close to other firms within the particular input-output production and consumption of which they are a part. According to Gordon and McCann [16] in the third model, the social network model, individual and collective actions of the firms differ from the behaviour associated with pure market contracting (present in the agglomeration model) or hierarchically organized relationships (in the

industrial complex model). Gordon and McCann [16] stress that there is nothing inherently spatial about the network model, though many of the preconditions for social networks, such as trust, common interest, social history are made possible by spatial proximity. Finally, Gordon and McCann [16] recognize that examples in the real world may contain elements of several models, while some types may be more significant than other.

Porter: “A cluster is a geographically proximate group of interconnected companies and associated institutions in a particular field, linked by commonalities and complementarities”.

Crouch and Farrell: “The more general concept of cluster suggest something looser: a tendency for firms in similar types of business to locate close together, though without having a particular important presence in an area”.

Rosenfeld: “A cluster is a very simply used to represent concentrations of firms that are able to produce synergy because of their geographical proximity and interdependence, even though their scale of employment may not be pronounced or prominent”.

Feser: “Economic clusters are not just related and supporting industries and institutions, but rather related and supporting institutions that are more competitive by virtue of their relationships”.

Swann and Prevezer: “Clusters are here defined as groups of firms within one industry based in one geographical area”.

Swann and Prevezer: “A cluster means a large group of firms in related industries at a particular location.”

Simmie and Sennett: “ We define innovative cluster as large number of interconnected industrial and/or service companies having a high degree of collaboration, typically through a supply chain, and operating under the same market conditions.”

Roelandt and den Hertag: “Clusters can be characterized as networks of producers of strongly interdependent firms (including specialised suppliers), linked each other in a value-adding production chain”

Van der Berg, Braun and van Winden: “The popular term cluster is most closely related to this local or regional dimension of networks.. Most definitions share the notion of clusters as localised networks of specialized organisations, whose production processes are closely linked through the exchange of goods, services/or knowledge”

Enright: “A regional cluster is an industrial cluster in which member firms are in close proximity to each other”

Lundvall and Borras: “The region is increasingly the level at which innovation is produced through regional networks of innovators, local clusters and the cross-fertilising effects of research institutions”

Asheim and Isaksen: “The crux of the regionalisation argument is that the regional level, and specific local and regional resources may still be important in firms’ effort to obtain global competitiveness...firms in the cluster rely on unique regional resources and local are cooperating when innovating”

Cooke and Huggins: “Clusters are geographically proximate firms in vertical and horizontal relationships, involving a localised enterprise support infrastructure with shared developmental vision for business growth, based on competition and cooperation in a specific market field”.

Maskell: “ The term cluster is used synonymously in the literature together with industrial agglomeration or localisation, while the term industrial district ...is often applied when wishing explicitly to emphasise to values and norms shared by co-localised firms.”

Becattini: “The industrial district is defined as a socio-territorial entity, characterized by the active co-presence, in an area territorially circumscribed, naturalistically and historically determined, of a community of people and a population of industrial firms. Within the district, differently from what occurs in others environment (for instance the manufacturing city), the community and firms tend, as we can say, to penetrate one other.”

Pyke and Senberger: “Industrial districts are geographical defined systems, characterised by a high number of firms active in different stages and in different modes of the production of a homogeneous product. A significant characteristic is that a large part of these firms are small firms or very small firms.... The various districts are specialised in different products with various degrees of complexity and with different final uses. ... A characteristic of the industrial district is that it has to be thought as a unique unity, a social and economic system...”

Table 1

Markusen [17] takes a different approach to classification than Gordon and McCann [16] and focuses on the processes of formation rather than structure of clustering. Markusen [17] distinguishes between four types of clustering: Marshallian industrial districts, hub-and-spoke districts, satellite industrial

platforms and state anchored industrial districts. In Marshallian industrial districts the business structure is dominated by small, locally owned firms, long-term contracts between buyers and sellers and low level of linkages with firms outside the district. In hub-and-spoke districts the structure is dominated by one or several large vertically integrated firms surrounded by suppliers, with substantial links to suppliers and competitors outside the cluster. In satellite industrial platforms there is a high degree of linkages with external firms, especially with the parent company. Finally, the state anchored districts are characterized by domination of a one or several large government surrounded by supplier and buyers.

Simmie and Sennett [18] extend Markusen's classification to include a fifth type of clusters-cities as international trading nodes, characterized by multiple clusters of innovative sectors. Simmie and Sennett [18], like Gordon and McCann [16] for their classification, stress that these five types of clusters are not mutually exclusive.

Bellussi [2] based on following four criteria: composition of the cluster, institutional characteristics, external links and development trends distinguishes between four types of clusters: canonical Marshallian, satellite, high-tech and evolutionary post-Marshallian cluster. Canonical Marshallian clusters are characterized by exclusive presence of family-owned SMEs with a high level of trust and cooperation and limited external links. Satellite clusters have a structure in which multinational companies (MNC) dominate and are also the centre of local networks, while external links are limited to links with MNCs subsidiaries. High-tech clusters are characterized by a wide variety of firms, both large and small, presence of universities or R&D and government research agencies and strong external links related to knowledge transfer. Evolutionary post-Marshallian clusters are formed through a dual structure with both large vertically integrated firms and a multitude of SMEs, presence of supporting agencies and growing external links.

3.2. *Knowledge spillovers, innovation and clusters*

Innovative capacity of firms has traditionally been explained by intra-firm characteristics: firm size and R&D expenditure. In the more recent literature, emphasis has been placed on factors that are external to the firm, the so-called knowledge spillovers. Knowledge spillovers refer to positive externalities that firms receive in terms of knowledge from the environment in which they operate [19]. Jaffe (as cited in Caniels and Romijn [20]) defines knowledge spillovers as intellectual gains that firms realize through exchange of information for which a direct compensation to the producer of knowledge is either not given at all or compensation that is given is less than the value of knowledge. Cappello and Faggian [19] emphasize that the key in this definition is that knowledge spillovers have a local dimension. In the context of knowledge spillovers and their localization, distinction between "tacit" and "explicit" is highlighted [21]. Tacit knowledge is difficult to transmit to others and can often be only transferred by means of close personal interactions, while explicit knowledge is codifiable in way that make inter-personal communication relatively easy. In practice, Scott [21] however cautions the distinction between tacit and the explicit "is rarely cut and dried and most forms of knowledge are complex, amalgam of the two types".

Empirical confirmation of the importance of spatial agglomeration and localized knowledge spillovers (LKS) in regard to knowledge-generation can be found in the work of Jaffe et al [22] who focus on the geography of patenting. A parallel line of research based on data from US Small Business Administration focuses on spatial concentration of innovative events in industry by Audretsch and Feldman [23; 24] and Feldman and Audretsch [25].

Jaffe et al [22] "trace" the influence from one patent to another as measured by the citations of prior patents that are part of the application for patent protection to the US Patent and Trademark Office. Therefore, in their model, to the extent that regional localization of knowledge spillovers is important, (new) citations should come disproportionately from the same geographic location as the originating patent. (Note that in his study Jaffe et al [22] do not focus on knowledge spillovers within firms in a given industry.) Jaffe et al [22] recognize that the problem to test for knowledge spillover-localization stems from the difficulty of separating knowledge spillovers from correlations that may exist due to

pre-existing pattern of geographic concentration of technologically related activities. Therefore, their model controls for other sources of agglomeration effects that could explain the geographic concentration of technologically-related activities. Based on this model and US patent data, Jaffe et al [22] conclude that knowledge-spillovers, or at least their trails (patent citations), are geographically localized. However, Scott [21] cautions “patents are notably troublesome as a measure of innovation, because not all innovations are patented, and not all patents are equally innovative or rewarding”.

Audretsch and Feldman [23] empirically examine the propensity for innovative activity to cluster geographically and link this geographic concentration to existence of localized knowledge externalities. Like Jaffe et al [22] they recognize (and control for) that one obvious explanation for innovative activities in some industries to cluster geographically more than other industries is the fact that location of production in these industries is more concentrated spatially. Analyzing data from US Small Business Administration’s Innovation Data Base (SBAIDB) they conclude that the “propensity for innovative activity to cluster spatially is more attributable to the influence of knowledge spillovers and not merely geographic concentration of production”. Audretsch and Feldman [24] extend their original research using the SBAIDB data to examine the link between propensity for innovative activity to spatially cluster and the stage of the industry life cycle. Audretsch and Feldman [24] conclude that there is a greater propensity for innovative activity to cluster during the early stages of the industry life cycle, while in the mature and declining stages of the life cycle innovative activity tends to be more highly dispersed.

The dissenting view of importance of LKS has been most prominently voiced by Breschi and Lissoni [26; 27] who argue that the role of LKS has been highly overrated. Breschi and Lissoni argue that there is a wide variety of mechanisms through which knowledge is exchanged, many which give rise to *pecuniary* rather than *pure* knowledge externalities, and that therefore the econometric studies that claim to find evidence of pure LKS may partly reflect cost advantages. Economies of specialisations and labour market economies (in the MAR externalities) are referred to as “pecuniary” or “rent” externalities, as opposed to knowledge spillovers which are referred to as “technological” or “pure knowledge” externalities. Pecuniary or rent externalities allow co-localized firms to access traded inputs and labour at a lower price than rivals located elsewhere, they occur through market interactions. Pure externalities, in contrast, materialize through non-market interactions. Finally, Breschi and Lissoni argue that tacitness is not an inherent characteristic of knowledge, but refers to the way knowledge itself is transmitted within an (epistemic) community.

Similar to Breschi and Lissoni [26; 27], Gordon and McCann [28] argue that the importance of local informal information spillovers is very much more limited than has been suggested. Using data from the London Employer’s Survey and other UK data, they argue that in the case of the London region standard agglomeration arguments appear to provide the best explanation of innovation dynamics.

Caniels and Romijn [20] suggest that a part of the explanation for diverse views in the LKS debate could be that the “precise meaning of the concept of LKS has often left vague”. They highlight that pecuniary externalities can be defined to include not only static MAR production cost advantages arising from local presence of specialized labour and inputs but could also be dynamic taking the form of *rent* spillovers arising from innovation. Furthermore, to advance the LKS debate they derive a typology of mechanisms through which the two types of externalities can stimulate learning and innovation at the firm level. Their typology examines how static pecuniary externalities and pure knowledge externalities affect the processes by which a firm modifies its organizational routines, which according to them is key for firm learning.

Breschi and Lissoni [27] also draw attention to the difference in discussion on localized knowledge spillovers between New Economic Geography (NEG) and the New Industrial Geographers (NIG) research. According to Breschi and Lissoni [27] most NEG are cautious of explanations based on pure knowledge externalities and insist on pecuniary externalities. They highlight some NEG arguments, such as that knowledge transmission is costless, or that costs do not depend on distance, or that LKS-based theoretical model can not be seriously tested as knowledge flows leave no tracks or as these flows are a result of complex non-market social interactions appropriate techniques to model them are missing. Breschi and Lissoni [27] note that though NIG is a much wider and more heterogeneous

group, NIG accept localized knowledge spillovers as a very important agglomeration force and suggest that most NIG explanations call for the distinction between tacit and codified knowledge.

Cappelo and Faggian [19] highlight the difference between geographers (NIG highlighted) and regional economists in explaining knowledge spillovers, where geographers tend to stress the physical proximity as a condition for the diffusion of knowledge while regional economists place an emphasis on relational capital to understand how knowledge spills over local areas. Relational capital closely resembles the concept of “social capital” (as in Putnam [29]), but according to Cappelo and Faggian [19] a distinction between the two concepts exists. They highlight that social capital exists where a local society exists, while relational capital refers to the capability of exchanging different skills and cooperating with other complementary organizations. They note that, though rare, interaction and cooperation can occur at a distance also. Using a sample of 217 firms located in the Veneto region, Italy they provide empirical support of the importance of relational capital on innovation activity. Recognizing that relational capital is an “intangible and complex asset that can not be measured directly” Capello and Faggian [19] measure relational capital indirectly via the collective learning channels proxied by local labour market (proportion of new employees for firms belonging to the local area) and cooperation with local supplier and customers.

Similar to Cappelo and Faggian [19], Boschma [30] argues that geographical proximity per se is neither necessary nor a sufficient condition for learning to take place. Boschma [30] points out that the impact of geographical proximity on learning can only be assessed in empirical studies when controlling for other dimensions of proximity, because they may act as powerful substitutes. Boschma [30] does stress that geographical proximity facilitates learning, by strengthening other dimensions of proximity. According to Boschma [31], these other dimensions of proximity are: cognitive, organizational, social and institutional proximity. He describes cognitive proximity as “people sharing the same knowledge base may learn from each other”. Boschma [31] argues that effective learning is accomplished by maintaining some cognitive distance while securing cognitive proximity. Organizational proximity is defined as the extent to which relations are shared in an organizational arrangement, either within or between firms. Social proximity is defined in terms of socially embedded relations between firms at the micro-level. Whereas social proximity is defined in terms of socially embedded relations between firms at the micro level, institutional proximity is associated with the institutional framework at the macro-level.

Like Boschma [31] and Cappelo and Faggian [19], Amin and Cohendet [32] argue for analyzing “space” and “proximity” in both geographical and organizational terms. Like Breschi and Lissoni [26] they reject the tacit/local versus codified/global distinction. They argue that “being there” defined as relational or social proximity is possible without physical contact. Furthermore, by highlighting the role that the trans-national community of US-educated Taiwanese engineers and their links with “home” regions in Taiwan have played in the excellence of the Silicon Valley cluster, Amin and Cohendet [32] argue that the knowledge creation within clusters is dependent upon trans-local connections. Like Amin and Cohendet [32], Gertler and Levitte [33] also focus on interaction between local and global knowledge transfer in an empirical study of the Canadian biotechnology industry. They document the importance of global networks but also highlight the value of local networks and in particular to raising capital.

Using FDI and cluster data from UK, Propriis and Driffield [3] analyze the importance of clusters in terms of knowledge spillovers from FDI to local firms, but also test for existence of “reverse” spillovers, those flowing from local firms to foreign-owned firms in the cluster. They conclude that firms in clusters gain significantly from local FDI, while in the non-cluster case, there are no such spillovers, merely “crowding” out. Moreover, they argue that investment by domestic firms in clusters also generates total factor productivity growth for inward-foreign investors located within clusters.

A different line of research on the effects of agglomeration on innovation proceeds on the basis of fundamental division of agglomeration economies into Marshall-Arrow-Romer (MAR) externalities and Jacobs externalities. MAR represent externalities produced and consumed only in a given sector, while Jacobs externalities are defined as externalities that flow between firms in all sectors. Jacobs externalities are also referred to as “urbanization” economies. Baptista and Swan [34] have found

evidence that MAR externalities are dominant, while Feldman and Audretsch [25] that Jacobs externalities are more prevalent. Using the US Small Business Administration Data Base (SBIDB) of product patents Feldman and Audretsch [25] test whether the specialization of economic activity within a narrow concentrated set of economic activities is more conducive to knowledge spillovers or if diversity better promotes innovation. They also explore the effects of specialization and diversity at the firm level. Feldman and Audretsch [25] conclude that both at the level of firm and for the industry across geographic space there is empirical evidence to support the diversity thesis. Similarly, Gordon and McCann [28] studying innovation among firms in greater London region argue that there is a greater influence of urbanization rather than localization effects.

In positioning the LKS debate in the context of clusters, Scott [21] highlights that much of the LKS empirical research is largely based on aggregate measures of innovation in which data for many sectors are pooled together. Furthermore, even if data is segmented by sectors as in analyzing the MAR vs. Jacobs externalities, this distinction is not very useful for clusters as these encompass more than one industry. In addressing the question of how do (firms in) clusters create knowledge, Malmberg and Power [35] extensively review the theoretical and empirical literature on clusters and propose and test three possible hypotheses for knowledge creation in the clusters. The three tested hypotheses are: knowledge in clusters is created through various forms of local inter-organizational collaborative interaction, knowledge in clusters is created through increased competition and rivalry, and knowledge in clusters is created by spillovers following from local mobility and sociability of individuals. For the first hypothesis on importance of collaborative interaction, which according to Malmberg and Power [35] includes inter-firm business transactions, non-transactional forms of inter-firm collaboration and cooperation between firms and universities, Malmberg and Power [35] find that there is a mixed and contradictory set of empirical data. For the second hypothesis, Malmberg and Power [35] conclude that there is a lack of systematic empirical evidence suggesting that it is both too early to claim support for but also dismiss this hypothesis. Finally, for the third hypothesis that posits the cluster as nexus of inter-personal relations, Malmberg and Power [35] find that informal knowledge exchanges do occur across groups of professionals in clusters and that local labour mobility is a factor in contributing to the success of clusters in terms of knowledge creation. However, they caution that many firms see this increased labour mobility as more of a problem than an advantage.

Analyzing the impact of clustering on firm innovation, Beaudry and Breschi [36] propose a theoretical model in which cluster effects can be both positive and negative and can emanate from both demand and supply side. On the demand side, the advantages are sophisticated users, strong-user supplier interactions and informational externalities, while the disadvantages can be competition in output markets and strong relational ties that may reduce flexibility. On the supply side, the advantages include Marshallian externalities such as skilled labour and specialised inputs and localized knowledge spillovers and informational externalities, while the disadvantages can be the “congestion externalities” reflected in cost of labour and real estate and possible inward orientation that can lead to technological lock-ins. Beaudry and Breschi [36] further empirically test whether firms located in industrial clusters in UK and Italy are more innovative than firms located outside these clusters using patent data, company data and regional employment data from these two countries. Modelling the total number of patents produced by a firm, active in a specific industry and located in a cluster, they argue that clustering in itself is not a source of benefits for firms’ innovative activities. They argue that it is not the size of the cluster that matters, but that it is the innovations by peers within the cluster that matters.

3.3. Inter-firm networks and clusters

Inter-firm networks are a critical feature of clusters. Oliver and Ebers [37] provide an extensive literature review of inter-organizational networks. In their review of 158 papers in the field, they find 17 different theories describing and explaining inter-organizational relations and networks. Huggins [38] provides a review of recent empirical inter-firm studies based around the network concept. He groups studies based on methodological approach into survey-based and other quantitative studies and

interview-based and other qualitative studies, all reviewed studies focus on inter-firm networks in developed economies. Huggins [38] also studies four inter-firm initiatives in the UK and argues that the most potent form of inter-firm networks are formal groups though these are best facilitated initially through an informal structure. Given this, Huggins [38] argues that policies that aim to create formal networks from scratch should be abandoned, rather policies should strive to develop networks that resemble those that would have emerged without the policy intervention.

3.4 *Clusters in developing economies*

Van Dijk and Sverrisson [14] argue that one can find clusters everywhere, in developing economies also. Martin and Sunley [7] however note that clusters in developing countries are often based on low costs of production, low wages and have limited growth potential and low productivity and therefore are not convincing examples of a new logic of knowledge-based cluster. Porter [39] also notes that developing country clusters are shallow and “hierarchical” with only island-like competitive firms.

Schmitz and Nadvi [40], like Van Dijk and Sverrisson [14], stress that clusters are significant feature of a developing economies where the survival of small enterprises is dependent on high levels of social capital. Weijland [41] analyzes micro-enterprise clusters in rural Indonesia and argues that social capital has been critical for achievement of transaction cost reduction. Similarly, by comparing the performance of clustered firms in the garment industry in Peru with a control group of dispersed producers, Visser [42] provides empirical evidence of clustering advantages in form of cost reductions and information spillovers. Rabelotti [43] analyzes a shoe-manufacturing cluster located in Guadalajara, Mexico and finds empirical evidence of increased cooperation of firms in the cluster when faced with increased competition from trade liberalization. Another study [44] of a footwear cluster examines whether enterprises in Sino Valley, South of Brazil have stepped up their cooperation in response to intensified global competition in leather footwear. In this study, combining qualitative and quantitative methods, Schmitz [44] finds a substantial increase in bilateral vertical cooperation, but not in multilateral cooperation. Nadvi [45] studies a surgical instrument cluster in Pakistan, and draws on quantitative and qualitative data to examine how inter-firm ties, both vertical and horizontal have changed in response to market changes. Tewari [46], Kennedy [47] and Knorringer [48] study individual clusters in India and their response to external crises. Tewari [46] focuses on a woollen knitwear cluster facing a collapse of its largest export market and trade liberalisation in India. Kennedy [47] focuses on the cluster of tanneries in India’s Tamil Nadu region and shows how firm cooperation and joint action has enabled the cluster to address the “pollution crises”.

Focusing on knowledge generation in clusters in developing economies Bell and Albu [49] stress that technological change is not something firms in developing countries simply “buy-in” from outside. On the contrary, they argue technological change is rooted within the structure of the firm itself and therefore there is a much wider scope for understanding technological upgrading in developing countries than just focusing on “acquisition” of outside technology and equipment. Furthermore, they present a conceptual framework that distinguishes between clusters’ knowledge systems based on following three criteria: the difference in the complexity of technologies in the observed clusters; the difference in the distance between the observed cluster and the international technological frontier; and the difference of the individual clusters’ knowledge systems which are responsible for differing technological dynamism of observed clusters. Finally, to bridge the gap between the micro-level of analysis that focuses on firm learning and the meso-level that focuses on inter-firm networks, cluster structures that facilitate learning, Bell and Albu [49] highlight various intra-firm, intra-cluster and external cluster sources of knowledge for firms located on clusters.

Giuliani and Bell [50] also attempt to bridge the meso and micro level when analyzing the Chilean wine cluster. Instead of the common argument that the meso-level cluster shapes micro-level firm behaviour, they stress the importance of opposite direction of influence, where the “capacities of individual firms to absorb, diffuse and creatively exploit knowledge shape the learning dynamics of the cluster as whole”. Given this, they argue that the policy measures that foster intra-cluster collaboration might not do much, rather measures focused on strengthening firms’ knowledge bases might lead to greater knowledge creation and stronger intra-cluster diffusion. A similar argument for

clusters in developed economies is advanced by Beaudry and Breschi [36] who argue that it is not the size of the cluster that matters to innovation in clusters, but that it is the innovations by peers within the cluster that matters.

Building on Bell and Albu [49] work, Caniels and Romijn [51] combine the meso-micro framework to develop a taxonomy of advantages that firms in clusters experience in relation to their learning. Their taxonomy distinguishes between *spontaneous* and *facilitated* traditional, Marshallian cost advantages and also between spontaneous and facilitated advantages arising from knowledge spillovers. The key is that facilitated advantages, both cost and technological, can occur only as a result of deliberate inter-firm cooperation. Caniels and Romijn [51] apply this taxonomy to an empirical case study on farm equipment manufacturing in Pakistan's Punjab province. They conclude that while there is a considerable favourable impact of spontaneous mechanisms, that mechanisms relying on inter-firm cooperation that require joint investments and long-term commitment were not observed.

Two successful clusters in Chile, one in agro-industry and one in aquaculture, analyzed by Perez-Alman [52] represent a different case from the Pakistan cluster analyzed by Caniels and Romijn [51] in that they stress the role that *facilitated* mechanisms and institutions play in enabling coordinated learning among firms to improve firm capabilities, processes and products. Moreover, Perez-Alman [53] extends the importance of building institutions that enable collective learning processes from a condition for dynamic clusters to one necessary for achieving sustainable growth and development.

Typologies of clusters in developing countries depart from the typologies of clusters in developed economies and are largely based on stages of clusters in the industrialization process. Van Dijk and Sverrisson [14] present a typology of clusters in developing countries that includes five types of clusters. *Location clusters* are predicated on information sharing and proximity, *local market clusters* are based on facilitation of transactions and reduction of transaction costs, *local network clusters* are embodied the initial stages of specialization, *innovative clusters* introduce local innovation and, finally *industrial districts* are characterized by the evolution of an institutional structure supporting co-operation, innovation and marketing. Similarly, analyzing six case studies from Africa, McCormick [54] distinguishes between types of clusters in the industrialization process: *groundwork clusters* that prepare the way, *industrializing clusters* that begin the process of specialization, differentiation and technological development and *complex industrial clusters* that produce competitively for a wider market.

3.5 Clusters in transition economies of SEE

Clusters in transitional economies of SEE exhibits most of the features and challenges identified for the clusters in developing economies, but also face challenges specific to SEE region's socialist history. Therefore, understanding clusters in transitional economies of SEE would not be complete without understanding the socialist regional legacy and impact of post-socialism transformation on the region. As Radosevic [6] notes "in socialism, proximity was not an asset". He further stresses that potential regional linkages were not used as a source of efficiency improvements or innovation. The predominant linkages were inter-regional and where organized within individual sectors or within large "combinats". The dominance of inter-regional vertical production chains and existence of few horizontal intra-regional linkages were coupled with the social role of large firms in the local community. Thus, he points out that while state-owned enterprises (SOEs) were socially strongly embedded in the locality, their forward and backward linkages were rarely located in the locality. In conclusion, Radosevic [6] argues that, notwithstanding individual country differences, "the former socialist economies were deprived of agglomeration economies".

In addition to their "non-agglomeration" socialist legacy, there are several other challenges that transitional economies face in pursuing a cluster policy. These include regional disparities due to uneven distribution of FDI, power and technological capability asymmetries between small firms and international investors, pressure from the speed of change and lack of qualified interdisciplinary development practitioners [8]. Also, low level of social capital is highlighted as a constraint to economic growth [8; 55]. Specifically, Balkan countries are faced with "low levels of institutional

trust necessary for the opening up of pervasive and impenetrable social networks” [55]. Even in Slovenia, highlighted as “the prime candidate from which to learn about the success drivers of clusters in transitional economies” [56] both policy makers and researchers caution that many challenges still remain to be addressed [57]. Specifically, a recent analysis of three clusters in Slovenia has shown that companies lack the most important element for successful cooperation-mutual trust [57; 8]. IRE [58] report presents an overview of cluster policies and clusters in eight New Member States of the EU (Slovenia, Poland, Hungary, Slovak Republic, Czech Republic, Estonia, Latvia and Lithuania) and highlights present challenges.

3.6 Clusters in Bosnia and Herzegovina

Bosnia and Herzegovina was one of the less developed republics in the former Socialist Federal Republic of Yugoslavia (SFRY). Its GDP per capita in 1987 was around 70 percent of the average GDP per capita in Yugoslavia, and almost 70 percent lower than the GDP per capita of the most developed republic of the former Yugoslavia, Slovenia. At the beginning of the 1990s industrial production in BiH was concentrated in largest cities and dominated by twelve large vertically integrated, with highly-centralized decision making, enterprises (often referred to as “holding companies”). The industrial sectors included production of textile and leather goods, machines and metals as well as wood processing. In the service sector, civil engineering was strong, while agriculture production and processing contributed to some ten percent of output. The twelve large enterprises accounted for about 35 percent of GDP and about 40 percent of total exports. Evidence suggests that the success of these enterprises varied from industry while decisions were often guided by non-market decisions [59].

During the first half of 1990s, BiH experienced the most devastating economic collapse of any economy in the Central and Eastern Europe since World War II. The war changed the social and economic map of the country [60]. As IMF [61] highlights “disruptions of civil war left plant and machinery severely outdated if not destroyed. Know-how was lost through large scale emigration and internal displacement of the labour force. And those enterprises, previously part of BiH’s vertically integrated industries found themselves isolated and without traditional local markets. And with the break-up of Yugoslavia, access to markets in the Former Yugoslav Republics was disrupted”. All of this left BiH’s GDP at less than 20 percent of its pre-war level [60].

In 1996 with the return of peace a major effort has been undertaken focused on BiH’s post-war economic recovery as well as its transition to a market economy. BiH has made notable economic progress since 1996 with real GDP that quadrupled between 1995 and 2004, however much of the economic growth has been fuelled by over \$5 billion in donor loans and grants, rather than by private investment.

Most firms in BiH are now privately owned, although the share of the private sector in output was only about 50 percent in 2005. This is among the lowest in the SEE region where in Serbia and Montenegro it is 65 percent, 68 percent in Croatia and 71 percent in Bulgaria. Small and medium sized enterprises (SMEs) play a prominent role in BiH economy. EBRD reports that in 2001 in BiH there were only some 200 enterprises that were classified as large enterprises compared to over 30,000 SMEs. While BiH has the lowest number of large enterprises in the SEE and pronounced micro enterprise participation, its number of SMEs per capita is also the lowest in the SEE region. There are only seven SMEs per 1,000 inhabitants in BiH compared to over 27 SMEs in Bulgaria and Romania. Also, within the SMEs there is a pronounced share of micro-enterprises where over 85 percent of the SMEs in BiH are micro-enterprises with 10 or fewer employees. In terms of sectoral composition, SMEs registered in trade and services account for 57 percent while those registered in construction and manufacturing represent 27 percent. The SME sectoral composition closely corresponds to the sectoral composition of the value-added of GDP where the services sector accounted for over 60 percent of value added by sector of GDP in 2003. In 2003 the share of agriculture in value added was estimated at 10 percent and industry around 28 percent, of which some 12 percent was accounted by manufacturing.

Moreover, foreign firms do not play a prominent role on the enterprise scene in BiH, this is despite the fact that levels of FDI in BiH have steadily increased in the past years. However, FDI inflows per capita to BiH are still among the lowest when compared to other SEE countries. In 2002, the FDI inflow per capita in BiH was US\$ 61 compared to US\$ 86 for Croatia and US\$ 67 for Serbia and Montenegro. Low inflows of FDI have not only resulted in limited presence of foreign firms in BiH, but as World Bank (2005) argues, have resulted in poor integration of BiH enterprises into international production and distribution networks and their low competitiveness. BiH firms surveyed by World Bank and EBRD export only 10 percent of their sales and only firms from Serbia and Montenegro in the SEE region have lower export intensity [60]. Several empirical studies have shown that FDI has played an important role in transfer of technology in more advanced transitional economies. Also as noted, several authors have argued that the knowledge creation within clusters is dependent upon trans-local connections. However, in BiH this knowledge transfer is only beginning to take place and on a relatively limited scale. Finally, as Propriis and Driffield [3] highlight, not only do local firms gain significantly from FDI in their clusters in terms of localized knowledge spillovers, but so do foreign investors from “reverse” spillovers from local firms. Therefore, while clearly it would be difficult to infer from Propriis and Driffield [3] research on FDI and clusters in the UK that a possible motivation of foreign firms to invest in BiH are the “reverse” knowledge spillovers, BiH nevertheless may want to take a more integrative perspective to its FDI and cluster policy.

Finally, an important feature of the enterprise sector in BiH is the large size of its informal economy. World Bank [60] estimated the size of the shadow economy at over 34 percent of the GDP for 2000. The World Bank [60] and USAID [62] stress that the onerous government regulations including the business registration and inspection regime, burdensome taxes and social security contributions are the main causes for the high degree of informalization of the BiH economy. Moreover, these constraints also translate into higher cost of doing business for BiH formal firms compared to other SEE countries where transition has advanced more [63].

The EBRD and World Bank Business Environment and Enterprise Performance Survey (BEEPS) that surveyed over 180 BiH enterprises in 1999, 2002 and 2005 is the only report that provides BiH-representative enterprise performance data. The BEEPS samples included state-owned firms (SOEs), privatized firms and newly established firms operating in industrial and service sector. Data from BEEPS 2 survey show that close to 20 percent of surveyed firms in BiH are loss makers and that loss making is more pronounced in state-owned and privatized firms where over 30 percent of SOEs were loss makers compared to 10 percent for newly established firms. Capital and labour productivity data help explain the differences in the performance of firms where capital productivity was almost three times lower for SOEs than for “de-novo” firms, while labour productivity was again highest for de-novo firms. However, even “de novo” firms, a subset of private SMEs who appear to be an important “driver” of economic growth in BiH, exhibit lower capital and labour productivity than firms in most of the SEE region. The state-level BiH Small and Medium Enterprise Development Strategy [64] also highlights similar weaknesses in SME development: poor legal and institutional environment, low level of knowledge and skills in the field of new technologies and ICT, low levels of productivity and technology, overly narrow orientation of SMEs on local markets, and lack of vocational education.

Clusters in BiH, including their dynamism and role in knowledge diffusion and generation, therefore need to be considered in the context of relatively low levels of productivity of the enterprise sector and the legal and business environment obstacles to entrepreneurship highlighted above. Moreover, the BiH SME Strategy [64] notes that the “cluster concept is relatively new in BiH and represents a new development model”. It further highlights that it is the international donors that are assisting and identifying clusters in BiH. Specifically, USAID has developed a program that supports the development of a wood processing cluster and a tourism cluster, while GTZ has identified and supported three clusters: automotive parts, leather and textile. The EU has promoted a regional focus to economic growth by helping establish and providing support for five regional development agencies in BiH. Against this largely donor-driven initiative to foster clusters and regional economic policies, BiH like other SEE transitional economies, faces challenges to cluster development stemming from its socialist legacy. As highlighted, BiH socialist system was dominated by inter-regional vertical production chains with few horizontal intra-regional linkages. Given that due to the war several large

industrial “holdings” totally collapsed and their integrated production was never fully restored, BiH could have been less “burdened” by its non-agglomeration socialist legacy than some other SEE transitional economies. However, certain policy choices such as treating SOEs as a social safety nets, inefficient bankruptcy regimes, voucher privatization [61] have prevented BiH to shed this socialist legacy. On the other hand, as BEEPS data suggest, a relatively vibrant “de-novo” private small medium enterprise sector has emerged that is “free” to utilize potential horizontal intra-regional linkages as a source of efficiency improvements or innovation. However, these potential clusters in BiH face some challenges specific to BiH post-conflict environment that other SEE economies do not face. These include very low levels of trust and institutional structures that are mainly built along ethnic, and not regional or economic, lines.

Rocha and Sternberg [11] and Malmberg and Power [35] note that the success of a cluster and its ability to foster knowledge creation depends on various forms of local inter-organizational collaborative interaction, including interaction between firms and universities. The role of trust is highlighted as crucial in fostering collaborative interaction. However, as noted post-socialist countries are characterized by low levels of trust. More importantly, following the 1992-1995 war and ethnic cleansing in BiH, it has been suggested that that level of trust in BiH has even more declined. Hakansson and Sjöholm [65] using a first omnibus survey with over 1,800 respondents find that trust in BiH has declined and that “level of trust outside of family networks is relatively low”. Hakansson and Sjöholm [65] find that only 14.5 percent of respondents stated that other people can be trusted. Comparing this figure to the levels of trust in the SEE region, Mungiu-Pippidi [66] reports that trust in Macedonia is highest with 48 percent of respondents stating that most people can be trusted. Trust is lowest in Montenegro with 19 percent, while the figures for Bulgaria, Serbia and Romania were 30, 32 and 42 percent respectively. Finally, Hakansson and Sjöholm [65] also highlight that ethnically heterogeneous regions in BiH show lower levels of trust than ethnically homogenous regions. The EU-supported regional development agencies as well as USAID and GTZ-supported clusters encompass firms from municipalities from both entities and representing diverse ethnic groups. Therefore it will be worth examining to what extent these programs and policies have been able to overcome the low levels of trust in these regions and foster genuine collaboration. Finally, it is also worth noting that while before the war networks of cooperation between universities and research centres and enterprises existed, this cooperation is now almost non-existent [64]. In conclusion, although there is an increased awareness among firms of value of networks, business associations and growing support for their establishment and growth, some of these forms of collaboration have been primarily donor driven and their long-term sustainability and effectiveness in terms of knowledge generation has not been critically evaluated.

4. Conclusion

Cluster policies are gaining increasing importance in the discussion about the competitiveness of the economies of South-East Europe. Empirical research in this field, though very rich for developed economies, for the region of SEE is almost non-existent. This paper aimed to give an overview of the various aspects in the existing literature related to clustering and to indicate potential challenges for the development of policies that promote clusters in the economies of the SEE region and in BiH in particular.

References

- 1 Rodrik D, Institutions for high-quality growth: what they are and how to acquire them, NBER Working Paper 7540, 2000, NBER, Cambridge, MA.
- 2 Belussi F., In Search of a Useful Theory of Spatial Clustering, 2004
- 3 Proprijs de L. and Driffield N. The importance of clusters for spillovers from foreign direct investment and technology sourcing, Cambridge J. Econ. 2006; 30: 277-291.
- 4 Doyle P., Kuijs L and Jiang G. Real Converges to EU income levels: Central Europe from the 1990 to the Long Term, IMF Working Paper; 2001.

- 5 Gristock. Achieving Growth in Wider Europe: Understanding the Emergence of Industrial Networks. School of Slavonic and East European Studies, University College London, ESRC Programme: "One Europe or Several?" Project; 2003, available from: <http://www.ssees.ac.uk/esrcwork.htm>.
- 6 Radosevic S. Regional Innovation Systems in Central and Eastern Europe: Determinants, Organizers and Alignments. School of Slavonic and East European Studies University College London, ESRC Programme: One Europe or Several? Project; 2000, available from: <http://www.ssees.ac.uk/esrcwork.htm>
- 7 Martin R, and Sunley P. Deconstructing clusters: chaotic concept or policy panacea? *J. Econ Geogr.* 2003; 3: 5-35.
- 8 OECD and LEED program, Business Clusters: Promoting Enterprise in Central and Eastern Europe, available from: <http://www.oecd.org/dataoecd/7/8/35136722.pdf>
- 9 IRE, Innovating Regions in Europe, An overview of Cluster Policies and Clusters in the New Member States of the European Union, IRE network; 2005, available from: <http://www.innovating-regions.org/>
- 10 Rebernik M, Dermastia M, Kroslin T, Cluster initiatives, a firm foundation and entrepreneurship in Slovenia; 2004.
- 11 Rocha H. and Sternberg R. Entrepreneurship: the Role of Clusters Theoretical Perspectives and Empirical Evidence for Germany, *Small Bus. Econ.* 2005; 24: 267-292.
- 12 Rosenfeld S. Bringing Business Clusters into the Mainstream of Economic Development, *Eur. Plan. Stud.* 1997; 5(1).
- 13 Perry M. Clusters Last Stand, *Planning, Practise and Research.* 1999; 14(12): 149-152.
- 14 Van Dijk M. and Sverrisson A. Enterprise clusters in developing countries: mechanisms of transition and stagnation. *Entrepreneurship Reg. Devel.* 2003; 15: 183-206
- 15 Torre A. and Rallet A. Proximity and Localization, *Reg. Stud.* 2005; 39(1): 47-59.
- 16 Gordon and McCann. Industrial clusters: complexes, agglomeration and/or social networks? *Urban Stud.* 2000; 37 (3): 513-32.
- 17 Markusen. Sticky places in slippery space: a typology of industrial districts, *Econ. Geogr.* 1996; 72(3): 293-313.
- 18 Simmie and Sennett. Innovative clusters: global or local linkages? *Nat. Inst. Econ. Rev.* 1999; 170
- 19 Capello R. and Faggian A. Collective Learning and Relational Capital in Local Innovation Processes. *Reg. Stud.* 2005; 39: 75-87.
- 20 Caniels and Romijn. What drives innovativeness in industrial clusters? Transcending the debate. *Cambridge J. Econ.* 2005; 29: 497-515
- 21 Scott A. Entrepreneurship, Innovation and Industrial Development: Geography and the Creative Field Revisited, *Small Bus. Econ.* 2006; 26: 1-24.
- 22 Jaffe A, Trajtenberg and Hederson R. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, *Quart. J. Econ.* 1993; August. 578-598.
- 23 Audretsch and Feldman. R&D spillovers and Geography of Innovation and Production, *Amer. Econ. Rev.* 1996; 86 (1): 630-640.
- 24 Audretsch and Feldman. Innovative clusters and the Industry Life Cycle, *Rev. Ind. Organ.* 1996; 11: 253-273.
- 25 Feldman and Audretsch. Innovation in cities: Science-based diversity, specialization and localized competition, *Europ. Econ. Rev.* 1999; 43: 409-429
- 26 Breschi S. and Lissoni F. Localized knowledge spillovers vs. innovative milieux: Knowledge "tacitness" reconsidered. *Pap. Reg. Sci.* 2001; 80: 255-273.
- 27 Breschi S. and Lissoni F (2001), Knowledge Spillovers and Local Innovation Systems: A Critical Survey, *Ind. Corp. Change.* 2001; 10(4): 975-1005.
- 28 Gordon and McCann. Innovation, agglomeration and regional development. *J. Econ Geogr.* 2005; 5: 523-543.
- 29 Putnam R. Making Democracy Work. Civic Traditions in Modern Italy. Princeton, NJ: Princeton University Press; 1993.
- 30 Boschma R. Role of Proximity in Interaction and Performance: Conceptual and Empirical Challenges, Editorial, *Reg. Stud.* 2005; 39(1): 41-45.
- 31 Boschma R. Proximity and Innovation: A Critical Assessment, *Reg. Stud.* 2005; 39(1): 61-74.

- 32 Amin A. and Cohendet P. Geographies of Knowledge Formation in Firms. *Industry Innovation*, 2005; 12(4): 465-486.
- 33 Gertler M. and Levitte Y. Local Nodes in Global Networks: The Geography of Knowledge Flows in Biotechnology Innovation, *Industry Innovation*, 2005; 12(4): 487-507.
- 34 Baptista R. and Swan P, Do Firms in Clusters Innovate More? *Research Policy*. 1998; 27, 525-540.
- 35 Malmberg A. and Power D. (How) Do (Firms in) Clusters Create Knowledge?. *Industry Innovation*. 2005; 12(4):409-431.
- 36 Beaudry C. and Breschi S. Are Firms in Clusters Really More Innovative? *Econ. Innovation New Tech*. 2003; 12(4): 325-342
- 37 Oliver, A. and Ebers M. Networking network studies: an analysis of conceptual configurations in the study of inter-organizational relationships, *Organ. Stud.*1998; 19 (4): 549-583
- 38 Huggins R. The success and failure of policy-implemented inter-firm network initiatives: motivations, processes and structure, *Entrepreneurship Reg. Devel.* 2000; 12: 111-135
- 39 Porter M. *On Competition*, Harvard Business School Press; 1998
- 40 Schmitz H. and Nadvi K. eds. *Industrial clusters in Developing Countries*, Special Issue, *World Devel.*1999; 27(9): 1503-1514.
- 41 Weijland H. Microenterprise Clusters in Rural Indonesia: Industrial Seedbed and Policy Target, *World Devel*, 1999, Special Issue, 27(9): 1515-1530.
- 42 Visser E. A Comparison of Clustered and Dispersed Firms in the Small Scale Clothing Industry of Lima. Special Issue, *World Devel.* 1999; 27 (9): 1553-1570.
- 43 Rabelotti R. Recovery of a Mexican Cluster: Devaluation Bonanza or Collective Efficiency, Special Issue, *World Devel*, 1999; 27 (9): 1571-1585.
- 44 Schmitz. *Global Competition and Local Cooperation: Success and Failure in the Sinos Valley, Brazil*, Special Issue, *World Devel.* 1999; 27 (9): 1627-1650.
- 45 Nadvi K. Collective Efficiency and Collective Failure: the Response of the Sialkot Surgical Instrument Cluster to Global Quality Pressures, Special Issue, *World Devel.* 1999; 27(9): 1605-1626.
- 46 Tewari M., *Successful Adjustment in Indian Industry, the case of Ludhiana's Woollen Knitwear*, Special Issue, *World Devel.* 1999; 27 (9): 1651-1671.
- 47 Kennedy L. Coordinating for Survival: Tannery Pollution and Joint Action in the Palar Valley (India), Special Issue, *World Devel.*1999; 27 (9): 1673-1691.
- 48 Knorringa P. *Agra: An Old Cluster Facing the New Competition*, Special Issue, *World Devel.* 1999; 27 (9): 1587-1604
- 49 Bell M. and Albu M. *Knowledge Systems and Technological Dynamism in Industrial Clusters in Developing Countries*, Special Issue, *World Devel.*1999; 27 (9): 1715-1734.
- 50 Guliani E and Bell M. When micro shapes the meso: Learning networks in a Chilean wine cluster, SPRU, Paper No. 115, University of Sussex, 2004.
- 51 Caniels and Romijn. Firm-level knowledge accumulation and regional dynamics. *Ind. Corp. Change*. 2003; 12(6): 1253-1278.
- 52 Perez-Alman P. Cluster formation, institutions and learning: the emergence of clusters and development in Chile, *Ind. Corp. Change* 2005; 14(4): 651-677.
- 53 Perez-Alman P. Learning, Adjustment and Economic Development: Transforming Firms, the State and Associations in Chile, *World Devel.* 2000; 28 (1): 41-55.
- 54 McCormick D. *African Enterprise Clusters and Industrialization: Theory and Reality*, Special Issue, *World Devel.* 1999; 27(9): 1531-1551.
- 55 Bozovic I. Outside the Network: Social Capital and a Critique of Extended Trust, *Conference Proceedings: From Transition to Development*, ICES 2003.
- 56 Solvell O., Lindqvist, G, Ketels C, *The Cluster Initiative Greenbook*.
- 57 Rebernik M, Dermastia M, Kroslin T. *Cluster initiatives, a firm foundation and entrepreneurship in Slovenia*; 2004. unpublished script.
- 58 IRE. *Innovating Regions in Europe, An overview of Cluster Policies and Clusters in the New Member States of the European Union*. 2005 IRE network, <http://www.innovating-regions.org/>

- 59 Prasnikar J., Domadenik P., Gregoric A., Jazbec B., Mrak M. Bosnia and Herzegovina, Country Study for GDN Project, University of Ljubljana and Institute for South-East Europe, Ljubljana, Slovenia; 2002.
- 60 World Bank. Bosnia and Herzegovina, Country Economic Memorandum. Poverty Reduction and Economic Management Unit, Europe and Central Asia Region; 2005 May.
- 61 IMF. Country Report No. 05/198, Bosnia and Herzegovina: Selected Economic Issues, International Monetary Fund; 2005 June.
- 62 USAID. Small and Medium Enterprise Intervention and Gap Analysis: Final Report, Sarajevo; 2004.
- 63 World Bank. Bosnia and Herzegovina, Country Profile: Doing Business in 2006; 2006.
- 64 BiH Small and Medium-Sized Enterprise Development Strategy in BiH, 2005-2007. Draft for Public Debate; 2005.
- 65 Hakansson and Sjöholm. Who do you trust? Ethnicity and Trust in Bosnia and Herzegovina. Working Paper 216; 2005, EIJS Working Paper Series from the European Institute of Japanese Studies.
- 66 Mungiu-Pippidi A. Deconstructing Balkan Particularism: The Ambiguous Social Capital of Southeastern Europe, Southeast European and Black Sea Studies. 2005; 5(1): 49-68.

The Impact of Technological Transfer from Foreign Direct Investment (FDI) – Preliminary Results from Albania

Mamica Skenderi¹, Phil Wright², Costas Alexiou³

¹SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, mskenderi@seerc.info

²The University of Sheffield, Dept. of Economics, 9 Mappin Str, Sheffield, S1 4 DT, UK, P.W.Wright@sheffield.ac.gr

³City College, Dept. of Business Administration and Economics, 2 Kalapothaki Str, 54624 Thessaloniki, Greece, c.alexiou@city.academic.gr

An isolated country like Albania, with small size and strategic location, constitutes a very interesting case in point for investigating the technological effects of FDI on the local companies. The effects of FDI are expressed in terms of two concerns: direct and indirect effects. Employing an explanatory study based on interviewer-administered questionnaires with foreign companies, the paper attempts to explain the role of FDI on the upgrading of local companies answering the questions of how and why the technological transfer occurs. Preliminary results obtained from sample survey with foreign companies operating in Albania show that the impact of FDI is important in terms of direct technology transfer to local subsidiaries; however, the technology transfer to local companies is limited mainly due to little contact of FDI companies with the host economy. In case where there is contact with local economy, local suppliers benefit (assistance on quality, price and time delivery) more than local customers and local competitors. Main transmission mechanisms involve direct cooperation mainly through informal sharing of views, higher requirements put by foreign companies, as well as through demonstration/imitation effects. FDI effects are greater in the manufacturing sectors of food and beverages, furniture and production of construction materials.

Keywords

FDI, MNEs, technological transfer, spillover effects.

1. Introduction

The FDI phenomenon is accompanied by an increasing interest in scholars, economists, politicians and specialists concentrating on two main areas such as determinants of FDI flows and effects of FDI on host and home countries, measured both at a microeconomic and macroeconomic level. The focus of the present paper is the second area of concern, and more specifically, the examination of FDI technological impact on a country part of Central and Eastern Europe (CEE), such as Albania. FDI has turn out to be one of the main drivers of globalisation and is assumed to play a substantial role in the technological development and integration of host countries, considerably important for the transformation of CEE transition countries. Governments have liberalized their regimes and pursued various policies to attract foreign investment, focusing on the concern of how to track appropriate domestic policies so as to fully reap the possible benefits of multinationals (MNEs) in the domestic economy. These issues are particularly essential as FDI can support the transition countries' efforts to achieve levels of development comparable to those of western countries. As Hunya [1] put it, FDI accelerates the transformation process, and improves competitiveness and productivity in the region. These positive effects are mainly caused due to the transfer of advanced technology and knowledge from MNEs internally to their subsidiaries and externally to the domestic companies. MNEs are

considered to be powerful and effective means in disseminating innovative technology from developed to developing or transition countries with low technological levels and underdeveloped markets.

The paper is organized as follows. Section 2 introduces the theoretical framework to explain the way FDI influences local subsidiaries and local enterprises through technological transfer, as well as the mechanisms through which technology diffuses and supports the upgrading of local enterprises (including competitors, suppliers and customers). Section 3 follows with a review of the existing empirical work and sheds light into the different methodologies used to approach the subject. Section 4 assesses the main developments of FDI flows in the country of interest, in Albania. Section 5 sets out the research methodology and data employed. Section 6 presents preliminary results from the sample survey. Finally, Section 7 concludes the paper by providing some future implications.

2. Theoretical Framework – Direct and Indirect Technological Transfer Through FDI

2.1 Definitions of FDI and MNEs

Foreign investment can be divided into two categories: *FDI* and *portfolio investment*. One can observe that there does not exist a general definition on FDI, however, most definitions share common characteristics that distinguish it from portfolio investment. IMF (p: 86 [2]) and OECD (p: 2 [3]) provide this definition: “Direct investment is the category of international investment that reflects the objective of obtaining a lasting interest by a resident entity (“direct investor”) in one economy in an entity (“direct investment enterprise”) resident in an economy other than that of the investor. In general, what distinguishes FDI from portfolio investment is that foreign direct investors holds at least 10 per cent ownership or control of an enterprise and takes an active role and is directly involved in the entrepreneurship, management and decision making process of the foreign business entity.

When a company undertakes FDI and sets up foreign subsidiaries with at least of 10 per cent of foreign capital, in more than one country, it becomes “*multinational enterprise*” (MNE).¹ Once a multinational decides to undertake FDI, it has to make important strategic decisions such as the entry mode. There are various entry modes that a foreign company might pursue, however, the most well known ones are greenfield investments (starting everything from scratch), acquisitions (acquisition of part or all of an existing company) and joint ventures (partnership with one or more local firms).

2.2 The Rationale behind FDI and MNEs

Foreign investment is one out of various forms of entering a foreign market. Among others, two efficient forms of international flows, vastly analysed by international theory, are trade and licensing. Several theories have addressed the question of why a firm might decide and why is it profitable to undertake direct investment in a foreign country, rather than exporting or licensing. A number of scholars have proposed and tested many theories, particularly in the last 40 years, offering their contributions to the development of FDI theory. Hymer [4] was the first to introduce the main rationale about the existence of FDI and MNEs as the need to control production and marketing operations in various national markets in order to take fully advantage of the potential returns on corporate assets of skill and knowledge. According to Hymer’s theory, MNEs possess non-tangible productive assets, which they are able to exploit and make their entry into foreign markets more profitable than other forms such as exporting and licensing. This enables MNEs to compensate the costs (for example, transportation costs, cultural differences) of doing operations in a foreign country and compete successfully with domestic companies, which on the other hand have better knowledge and expertise of local markets. The firm-specific assets incorporate: specialized technological know-how about production; superior marketing, management and organizational skills; export related

¹ The terms multinational enterprise (MNE), multinational corporation (MNC) or transnational corporation (TNC) can be used interchangeably.

know-how and access to markets, contacts and networks; special skills in distribution and qualified relationships with suppliers and customers; as well as brand names and reputation. These are internalized within the firm and are transferred at low or zero additional cost from parent company to foreign branch plants.

Hymer's work was extended, among others, by Dunning's Eclectic Paradigm, which is one of the most quoted theories with the greatest explanatory power on FDI and MNEs. The MNE must possess some advantages over domestic firms that have to outweigh the costs it faces when invests abroad. According to the eclectic theory, a firm's decision to invest in a foreign country is determined by the existence of three different types of advantages: *ownership* advantage (firms' unique assets such as technological, marketing, or management know-how; a brand name; the benefits of economies of scale, etc.), *location* advantage (ownership advantages are possible to move between different locations and can therefore be transferred to a foreign country), and *internalisation advantage* (the firm has to be capable to internalise its advantages to the host economy) therefore the acronym OLI [5]. In order for FDI to take place all three advantages must be present simultaneously.

Hymer's theory about the existence of FDI and MNEs, and Dunning's Eclectic Paradigm (particularly the third advantage – the internalisation process) imply that MNEs have great potentials for technological transfer to the foreign subsidiary operating in the host economy. A number of interesting questions arise: 1) how MNEs transfer firm-specific asset to their affiliates and why these affiliates are supposed to be more productive than domestic firms? 2) MNEs transfer technology within themselves (transfer from parent to affiliate), however, do other domestic firms benefit from the presence of foreign affiliates? 3) If there is technology transfer to other domestic firms, how and under what conditions does this transfer take place?

2.3 Direct Technological Transfer

FDI is considered as the best and cheapest channel for technology transfer across national boundaries, as well as between firms (UNCTAD [6], Blomström and Kokko [7]), since most R&D takes place in MNEs located in the most advanced economies (UNECE [8], Hoekman et al., [9]). Technology is a broad concept, however, it is widely accepted that the technology transferred through FDI can take two forms: hard technology and soft technology [10]. Hard technology consists of physical investment: plants, equipments, and machineries. Hence, hard technology is supposed to include aspects of embodied knowledge in the machinery and equipment. On the other hand, soft technology includes: knowledge, management/organization system, production processes and expertise. Soft technology is supposed to include aspects of disembodied knowledge as a result of the transfer of operation skills. The terms "technology" and "knowledge" are used interchangeably in the FDI literature.

FDI constitutes in a direct transfer of foreign technology to its subsidiaries, which in turn leads to higher productivity. This is referred as the direct effect of FDI. Reflecting parent company's practices in the home country, MNEs affect directly their subsidiaries by injecting them directly foreign capital, equipments, technology, processes, organizational/marketing/management skills, and know-how. As mentioned in the section above, Hymer's widely accepted theory and Dunning's Eclectic Paradigm provide the implications on FDI direct transfer. Due to activities taking place within the subsidiaries, it is strongly suggested that as integrated part of the parent company, foreign subsidiaries have access to specific advantages and resources. If foreign subsidiaries have sufficient absorptive capacities to assimilate this know-how, the direct technological flow will in turn lead to improvement in their performance including output, productivity, efficiency, as well as exports. The technological transfer from parent to foreign subsidiaries company is the rationalization that scholars provide to explain why foreign subsidiaries are expected to do better than local firms; the latter do not have access to such competitive assets. These are the arguments that support the empirical work on the subject starting with the assumption that foreign firms are considered to have higher productivity and overall better average performance (including profitability) than the domestic ones (as well as higher wages, skills, and growth). The superior performance of foreign firms is already well documented in the literature.

Following Hymer and Dunning, Blomström and Kokko [11] provided an important element in the MNEs framework, according to which, the parent corporation and its subsidiaries constitute a highly integrated organizational unit, within which knowledge-based assets are created. The parent company has the exclusive power to decide about the exploitation of these assets. The possession of such technological assets, makes the foreign subsidiaries be superior related to domestic firms in the host economies. Blomström and Kokko [11] summarize:

“It is well known that multinational corporations undertake a major part of the world’s private R&D efforts and produce, own and control most of the world’s advanced technology. When a MNC sets up a foreign affiliate, the affiliate receives some of the proprietary technology that constitutes the parent’s firm-specific advantage and allows it to compete successfully in an environment where local firms have superior knowledge of local markets, consumer preferences and business practices. This leads to a geographical diffusion of technology, but not necessarily to any formal transfer of technology beyond the boundaries of the MNC.” (p: 10, [11]).

Moreover, there are even other interpretations on why foreign-owned plants may be superior. Harris and Robinson [12] argued that domestic firms, in contrast to foreign firms, are possibly constrained to obtain capital cheaply from abroad, which in turn reduces their ability to invest in superior technology. Another interpretation relates to the entry of foreign firms and its implications for industrial production. Aitken and Harrison [13] agreed on the competitive advantages of foreign firms over domestic ones and summarized the arguments for the assumed superior performance of foreign firms such as, transfer of better and newer equipment and machinery combined with transfer of intangible assets in the form of technological know-how, management and marketing capabilities, organized networks with suppliers and customers abroad, and trade contracts. However, after this, the authors emphasized that foreign firms may reduce, particularly in the short run, the productivity of domestic firms. The new entry of foreign-owned plants will increase the level of competition. On one hand, the new competitive conditions induce local firms to replace inefficient technologies and organizational practices through imports of capital goods and R&D expenditures, which in turn increase overall industrial productivity. On the other hand, however, if the market expands at a lower rate than the increase in capacity due to new entrants (foreign firms), then this would raise the average costs of domestic competitors as they lose market shares due to foreign entry. Hence, this situation would result in decline in productivity levels of domestic competitors, which are forced to operate at sub-optimal scales. Nevertheless, the other side of the coin exists too. Rotemberg and Saloner [14] argued that in case that there is increase in the number of firms associated with an increase in the demand conditions without any loss in market shares, this would result in external economies which shifts cost curves down for all the firms.

Literature provides also arguments that FDI subsidiaries may have lower productivity and lack significant advantages as compared to their domestic counterparts. Harris and Robinson [12] summarized a number of factors that could lead to lower productivity of foreign firms. Foreign subsidiaries may have lower efficiency, particularly in the short run, as time is needed to assimilate new plants (greenfield or acquired) into the FDI network. This may be also associated with big cultural differences between home and host markets or hostile government policies, which in turn can lead to long-term problems, as well as decline in productivity. Other factors related to the reasons why foreign firms may be less productive refer to the fact that these firms might keep their high value-added operations (R&D and new products) at home, engaging in lower value-added activities in the host country. Hence, lower productivity would result from the use of lower-skilled workers (who get lower wages), as well as older technology. Furthermore, another argument showing that MNEs operate in low value-added plants is that addressing to MNEs operating in oligopolistic markets, and where products are in the mature stage of their life-cycle. The assets used by MNEs operating in such markets may also be in a comparable stage in terms of their life cycle.

Despite the existence of various assumptions and arguments in the relevant literature, most academics argue that positive effects of MNEs outweigh the negative effects, implying that the overall performance of foreign firms is better compared to that of domestic counterparts. To sum up, direct FDI effects relate to the direct transfer, from parent to subsidiary, of foreign capital, equipments,

technology, know-how, processes, products and skills. It is strongly suggested that foreign affiliates, as integrated part of the parent company, have access to competitive advantages and resources possessed by the parent. Assuming that foreign affiliates use “best practice” technology, performance differences (particularly, superior productivity) arise between local firms and FDI subsidiaries. However, the magnitude of these direct effects depend on a number of factors such as the motives for investment, the mode of establishment, the scale of initial investment, the type and age of technology employed, the number and origin of workers combined with their training and wages, the extent to which the foreign company brings technology and specialized staff in its subsidiary, the reinvestment in the subsidiary, the characteristics of the host country [15].

2.4 Indirect Technological Transfer

Foreign enterprises do not affect only their subsidiaries, but also the behavior and performance of other local firms in the same sector or even in other sectors. These indirect effects are referred as spillovers or externalities. If MNEs transfer superior knowledge and technology to their foreign subsidiaries, it is expected that these firm-specific advantages are not perfectly internalized by the subsidiaries. It is possible that also other local firms in the host country might benefit from the uncompensated benefits of foreign presence. Some of the superior knowledge of products, processes or markets, technology and management, as well as marketing practices of foreign subsidiaries would “spill over” to domestic firms, without incurring any costs, through non-market transactions. These indirect effects are called technological spillovers or productivity spillovers.² Local firms would be able to internalize these spillovers and absorb them effectively under some conditions, such as absorptive capabilities (skills, knowledge, experience) and technological conditions. In this case positive spillovers are generated.

“The establishment of a foreign affiliate is, almost per definition a decision to internalize the use of core technology. However, MNC technology may still leak to the surrounding economy through external effects or spillovers that raise the level of human capital in the host country and increase productivity in local firms” (p: 10, [11]).

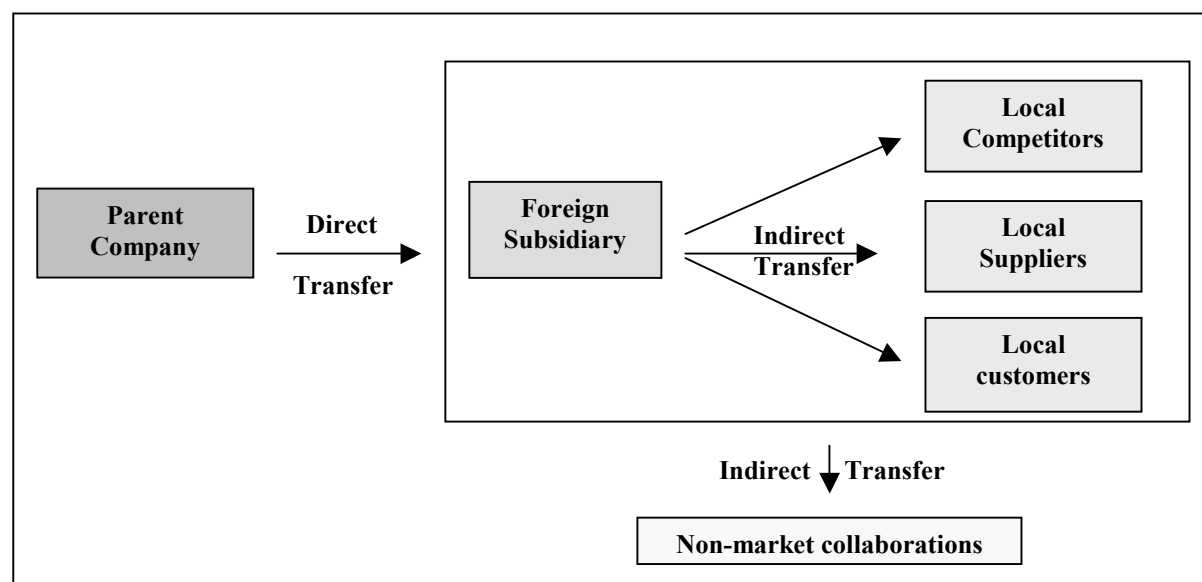


Figure 1 Direct and Indirect Transfer of Technology, Knowledge, Processes and Skills.

Most of the theoretical literature recognizes that’s pillovers may run in two directions: directly or internally between domestic firms that may be in direct competition with the foreign enterprises (intra-

² Usually, the terms technology spillover and productivity spillover are used interchangeably.

industry or horizontal spillovers); and indirectly or externally to other firms in the host economy that may be vertically integrated the foreign enterprises (inter-industry or vertical spillovers) (see Figure 1). In particular, inter-industry spillovers may derive from any backward linkages (suppliers) or forward (customers) between domestic firms and foreign affiliates. Backward linkages (sourcing) are created when foreign affiliates acquire materials, intermediate products, or services from domestic suppliers. Forward linkages (distribution) are created when foreign affiliates sell goods or services to domestic customers or when domestic firms use the final products of the foreign affiliates as raw materials in their local production processes. Even though theory on spillovers highlights the importance of the all types of linkages, forward and backward, in general evidence is strongest and most consistent in case of backward linkages with domestic suppliers.

Positive Intra-Industry Spillovers

Intra-industry spillovers occur when MNEs do not fully internalize the benefits of their competitive advantages and their presence increases the productivity of local competitors in the host economy. Based on the work on the previous scholars, there have been identified at least four mechanisms or channels through which technology might be diffused horizontally from foreign companies to other firms in the economy: demonstration – imitation effect, competition effect, cooperation, training effect and exporting effect [17], [7], [17], [18], [19].

First, the proximity of local firms to foreign enterprises can lead to demonstration effect. When foreign firms introduce new products, processes, technologies as well as organizational forms, they provide to other local companies a demonstration of increased efficiency and productivity. Second, the transfer of capital and technology stimulates competition in the local market. Domestic companies face a greater competitive pressure, which induces them to introduce new products in order to protect their market share and adopt new management methods so as to increase productivity. Third, knowledge can be transferred indirectly through the movement of labor. When MNEs subsidiaries hire local workers, the human capital may be enhanced further through organization of training facilities and on-the-job learning. Consequently, indirect effects arise when local personnel trained in the foreign subsidiary decide to leave the firm and move to other domestic firms or help establish new business. Benefits may arise too if superior management skills of foreign MNEs stimulate local suppliers, distributors and competitors to improve their own management skills. Finally, domestic firms can learn from foreign companies, most likely through imitation and collaboration, how to export and penetrate export markets from MNEs. Export-oriented foreign firms motivate local firms to become exporters, and enhance the development of domestic firms' export capabilities by bringing access to buyers to the countries that their products are sold. It is generally argued that spillovers derived from export oriented FDI are larger than domestic market oriented FDI.

However, spillovers highly depend on a number of factors that determine their materialization. The first group of factors refers to the host country' characteristics (market size, regulations, incentives offered to FDI and openness to foreign trade). The second group of factor involves the type of MNEs operating in the host country (nature or type of activity undertaken, kind of technology employed, level of human capital and wages). More specifically, there is low possibility for spillovers to emerge when foreign firms keep most of their value-added operations at home and concentrate at lower value-added activities at host country [12]. Therefore, the use of older technology and low-skilled workers (due to low costs) become impediments not only to the generation of spillovers, but can also lead to lower productivity levels of foreign affiliates compared to domestic firms. The third group of factors refers to the conditions of host companies (technology and productivity gap between local and foreign companies, absorptive capacity, geographical proximity, R&D expenditure). Positive spillovers are realized only if local firms have adequate social capabilities and absorptive capacities to absorb foreign technologies and skills. It is often difficult for relatively backward countries to acquire the necessary social capabilities and absorptive capacities that allow domestic firms to benefit from the spillovers potentially available in the economy [12].

Positive Inter-Industry Spillovers

Inter-industry spillovers refer to the spillovers taking place due to linkages between foreign firms with their local suppliers and customers (distributors or sales organizations). Whereas MNEs have an incentive to minimize technology leakage to competitors, they seek to promote vertical spillovers. The diffusion of technology across customers, and most importantly suppliers, appears to benefit foreign subsidiaries as their clients and suppliers become more efficient and competitive, by achieving improved levels of quality, delivery response and lower costs. This effect is referred as foreign linkage – cooperation effect. Cooperation of foreign enterprises with downstream customers and upstream suppliers increases spillovers. Hence, some of the channels of intra-industry spillovers, such as demonstration and imitation effects, and labor mobility between firms, contribute to inter-industry spillovers too. Given that the FDI effect is not limited to the initial direct effect to its subsidiaries, but involves also linkages with other parties in the economy, the later are considered as a multiplier effect of initial FDI effect. In the case of backward linkages with suppliers, the multiplier can be referred as “value-chain multiplier effect” while in the case of forward linkages with customers it is referred as “consumption multiplier effect” [15].

The most quoted vertical spillover is the backward linkage channel with suppliers. In an attempt to improve the quality standards of the suppliers, this channel operates through four ways ([20], [21], [22]). First, there may be direct knowledge and information transfer, quality control and inventory management, training, technical as well as management/marketing/financial assistance from foreign enterprises to their local suppliers. Second, MNEs may place higher requirements or even provide financial assistance for the improvement of product quality and on-time delivery, which in turn provide incentives to domestic suppliers to improve production and technology. Active spillover effects arise when MNEs put specific requirements to their local suppliers in terms of cost/price, quality and time delivery. On the other hand, passive spillovers take place when MNEs do not put specific requirements to the suppliers, but let them decide on how to improve their performance. Third, there may be indirect knowledge transfer through movement of labor (similar with intra-industry spillovers). Fourth, there may be a growing demand for intermediate products due to MNEs entry, which provide opportunity to local suppliers to exploit the benefits of economies of scale. Last but not least, MNEs that acquire domestic firms may decide to source intermediates abroad, which would in turn break existing supplier-customer relationships and increase competition in the intermediate products market.

The focus of researchers, academics and host governments is in the development of backward linkages and spillovers, as they provide benefits to both domestic firms and foreign subsidiaries, as well as to the host economy as a whole [23]. From the local supplier’s point of view, the direct effect of linkages and spillovers is manifested in a rise in output and employment, combined with transmission of knowledge and skills between both parties. Further, a dense network of linkages can encourage: productivity and efficiency growth, technological and managerial capabilities, and market diversification for the firms involved. On the other hand, from the foreign subsidiaries’ position, linkages with local suppliers can lower production costs in the host economies, offer information on local markets, and permit greater specialization and flexibility by adapting better technologies and products to local conditions. At last, for the host economy as a whole, linkages with local suppliers and spillovers can promote economic growth and provide benefits for the balance of payments (when local inputs, used as raw materials, substitute for the imported ones). All these effects are referred as wider effects of FDI on the host economy.

There are various factors that influence the generation of inter-industry spillovers. Apart from the factors that are similar to the ones influencing intra-industry spillovers, there are some other specific factors that play also important role in inter-industry spillover. OECD [23] stressed that the main factor is the availability of domestic supply capacity, with the lack of efficient domestic suppliers being a common obstacle to the creation of linkages, particularly in developing and transition countries. Thus, the decision to choose local suppliers and source locally highly depends on: cost/price, quality, reliability, and flexibility of local suppliers compared to foreign suppliers. Increasing competitive pressure forces foreign subsidiaries to strictly select suppliers that can meet the requirements in cost, quality and time delivery. However, local suppliers many times have difficulties

in matching these requirements, particularly in developing and transition countries, which often leads foreign affiliates to use as suppliers other affiliates operating in the host country or use outsourcing. On the other hand, UNCTAD [21] highlighted the importance of foreign companies' characteristics in the probability of spillovers to arise. First, investment motives and strategies of MNEs: domestic-oriented affiliates generally purchase more locally than do export-oriented affiliates. Second, technology and market position of MNEs: affiliates producing standardized products tend to outsource more to local companies rather than companies producing highly specialized products, which tend to outsource less to local companies. Third, age of foreign affiliate: the higher the experience of MNEs in a foreign country, the higher the number of managers recruited locally and the higher the knowledge of sourcing locally, which lead to lower costs of sourcing locally. Fourth, mode of establishment: foreign affiliates that have entered through M&As tend to have stronger links to local suppliers as they have already established network from the acquired local firm. Fifth, size of affiliate: large affiliates tend not to source locally, as in general local suppliers lack the capability of providing large volumes of raw material. Last but not least, sector in which the foreign affiliate operates: foreign affiliates involved in technology that can be divisible into discrete stages can outsource easier than when involved in a continuous process; thus, the most feasible sectors for outsourcing are those in which products are standardized such as low value-added textiles, electronic components, some automobile components and mining.

Negative Intra and Inter Industry Spillovers

Yet, not all the foreign enterprises' activity leads to positive spillovers. MNEs are profit maximizing, and thereby not interested in increasing benefits for other enterprises without obtaining a good price for it [24]. They can prevent spillovers to domestic firms by preventing technology leakage and spillovers from taking place, as both parties are in direct competition [25], [22]. In particular, an important feature of the choices made by foreign subsidiary' management is the minimization of the probability of imitation. It is unlikely to be in the interest of the foreign firm to share its firm-specific advantages with the domestic firms. Therefore, with imperfect intellectual property rights, MNEs tend to make entry decisions on the basis of limiting knowledge and information leakages as far as possible. Foreign firms can achieve this through formal protection of their intellectual property, trade secrecy, as well as by paying higher wages than local standards (in order to discourage highly trained employees to leave MNEs) or locating in countries or industries where domestic firms have limited innovative capacities to begin with. Moreover, even if there is information leakage, foreign affiliates may provide too few and/or wrong kind of technology (even outdated technology) to their domestic counterparts [8]. Further, domestic firms may not have the necessary absorptive capacity and human capital to absorb the latest technology, and/or there may be even skill mismatches when domestic firms hire trained staff by MNEs [12]. Also, a "brain drain" effect can be created in local firms, when foreign firms attract talented and skilled workers away from local firms [26].

In addition, foreign enterprises can also negatively impact their domestic competitors by reducing the productivity of the former, particularly in the short run. There is a simple story behind this. The entry of foreign firms leads to a more fierce competition. Given increasing costs and the fact that foreign firms can draw the sales and demand away from domestic firms, the latter are forced to reduce their production, which in turn causes a decline in the productivity [13]. Frequently, the increased concentration due to the presence of foreign firms and the fierce competition make less efficient and weaker firms loose their market shares or even force them exit the market. This is known as "crowding out" effect.

Negative vertical spillovers may also appear when foreign enterprises eliminate or "crowd out" domestic suppliers by relying on foreign ones (it is argued that there is a tendency of foreign firms to support foreign suppliers). Finally, foreign firms may also limit exports to domestic competitors. All these effects have a negative impact not only on the productivity of domestic firms, but also may have a wider negative impact in the whole economy growth.

3. Empirical Review – Methodological Approaches, Findings, Limitations

Economic theory has identified both positive and negative effects of foreign presence on host economies. However, the picture becomes less clear when turning from theory to empirical results. Empirical literature, with just few exceptions (for instance, Damijan et al., [27] for transition countries), usually confirms that subsidiaries of MNEs in the host countries, including developed, developing and transition countries, have on average higher productivity levels than purely domestic firms³ (direct effect of MNEs to their affiliates) [28], [13], [29], [30], [12], [31], [32]. Yet, when coming to spillovers, these effects are difficult to measure. In general, there are three methodological approaches that have been assigned the task to measure spillovers. First, due to data limitations, the early efforts to provide evidence on spillovers is provided through case study approach, which has been important in its contribution to FDI theory. Second, initial steps in conducting econometrics for FDI spillovers, are done using industry or sectoral level approach based on cross-sectional database. Third, more sophisticated recent techniques involve the adoption of micro-level analysis or firm level panel data, replacing cross-sectional data, in order to estimate the effects of foreign firms on total factor productivity of domestic ones. Nevertheless, all of these methods have their own advantages as well as limitations.

3.1 Review of Econometric Findings

The majority of recent literature on FDI spillovers is based on econometric approach focused on developed, developing and more recently in transition countries. However, due to the fact that spillovers are difficult to measure, the approach adopted in the existing empirical literature avoids the question of how productivity spillovers actually take place [33]. It concentrates only to the simple issue of whether or not the presence of foreign companies affects productivity in domestic companies. Judging from a number of studies, empirical evidence of productivity spillovers through FDI is mixed. The early work of Caves [34] and Globerman [35] on FDI spillovers was the starting point for the econometric examination of spillover patterns. Since then, however, their empirical models have been extended and sophisticated (there is an impressive increase in the number of such studies particularly in the last 5 years). Most of the studies on spillovers are usually done at the firm, industry, or sectoral level panel data, since spillovers occur between firms (foreign and domestic). Panel data are taken from the financial statements of individual firms, which are usually provided through industrial surveys carried out by National Tax Offices or Statistic Offices. Panel data are to be preferred as they capture certain country-specific factors (important if host country characteristics matter) that do not appear in cross-country time series data [8]. The drawback of industry level panel data is that they measure only intra-industry spillovers, and cannot measure adequately inter-industry spillovers. By contrast, firm level panel data capture both intra-industry, as well as inter-industry spillovers. In general, the studies using econometric approach, measure short run effects of foreign firms on domestic firm's productivity. In examining horizontal spillovers, output level or labour productivity or total factor productivity in domestic firms is regressed on various factors (for instance, capital intensity, production scale and labour skills) supposed to impact productivity, one of which is the extent of foreign firm's presence, which is usually calculated as the share of employment or sales or assets of foreign firms over.

The first group of studies includes a plethora of early industry level studies that examined the existence of intra-industry spillovers using cross-sectional data and showed positive correlation between foreign presence and domestic productivity. The pioneer study was that of Caves [34], which was carried on Australian manufacturing sector. Significant intra-industry (horizontal) spillovers were revealed when a foreign presence in employment was included as explanatory variable among other characteristics in total factor productivity. Cave's study was followed by that of Globerman [35] for Canada and Blömstrom and Persson [36] for Mexico. Similar to Caves [34], both studies used cross-

³ In general, the empirical studies aiming to examine spillovers, start with an estimation of productivity differences between foreign and domestic firms and most of them find higher levels of productivity in the foreign firms. The next step then, is to determine spillover effects over domestic firms.

sectional data for the manufacturing sector and found positive and statistically significant intra-industry spillovers. Subsequent studies of Nadiri [37] for France, Germany, Japan, and United Kingdom, Blömmstrom and Wolf [38] for Mexico, and Hurbert and Pain [39] for UK revealed similar results. However, this group of studies have been strongly criticized for their simplistic nature by a number of researchers, who emphasize that positive results can be explained by reasons other than the existence of spillovers from MNEs [17], [13]. Positive externalities appear only due to the fact that MNEs tend to locate in relatively high productivity sectors, and that the presence of foreign firms (which increases competition) might induce the exist of less competitive domestic firms, which in turn raises the average productivity level of the domestic firms, which in turn raises the average productivity level of the industry. Despite the critics, however, this econometric literature has its own merits; it is considered to be important as it presents the first attempt to quantify the mechanisms documented in theoretical literature.

The second group of studies includes a number of more recent works, which use panel firm level data. The later has provided a more accurate approach compared to industry level studies to investigate whether foreign presence in the sector or region is correlated to the productivity of domestic firms over time. Most of this work makes reference to developing or transition countries and find the absence of positive intra-industry spillovers, and even imply that foreign presence is not beneficial to the firms of the host economy. These studies usually suggest for negative spillovers and show that technology dissemination from foreign firms to domestic counterparts did not produce the externalities expected by the traditional spillover model. The pioneering studies of this kind, that not only casted doubt on the existence of positive spillovers but showed also that foreign presence affected negatively the productivity of local firms are that of Haddad and Harrison [28] for Moroccan manufacturing sector and Aitken and Harrison [13] for Venezuelan manufacturing sector. Aitken and Harrison noted that their findings would have been totally opposite (positive spillovers rather than negative) if they would have not taken into account that foreign affiliates tend to locate in high productivity sectors. Similar results are reached for the case of transition economies⁴ in the studies of Djankov and Hoeckman [40] for the Czech Republic, Konings [41] for Bulgaria, Romania and Poland, Damijan et al., [27] for eight transition countries, and Zukowska-Gagelmann [42] for Poland. It is very interesting to remark that most researches done in the transition economies provide for negative spillovers of foreign enterprises to local industries.⁵ Overall, the negative effects found in these countries imply that the negative competitive effects of foreign subsidiaries outweighed any positive technology and productivity improvements from demonstration effect and labour movement effect.

The third group of studies provides remarkable evidence on the factors that determine the existence or not of spillovers. Using again panel data techniques, these studies show that productivity spillovers may exist but they are dependent on various factors such as technology and productivity gap, absorption capacity, and geographical proximity. The most quoted factors that are analyzed in these studies are the role of technological and productivity gaps between local firms and foreign firms. Technology gap appears to be one of the most important conditions favoring a positive impact of foreign investments on the productivity of domestic firms. A too large gap could present an obstacle for spillovers to occur, as local firms would be unable to benefit from transfer of knowledge, as well as to improve competitiveness through modernization of technology. Among others, Cantwell [43] for Europe, Kokko [44] for Mexico, and Kokko, Tansini and Zejan [45] for Uruguay found that spillovers depend on technology gap between foreign and domestic companies. Girma [46], Girma, Greenway and Wakelin [47] and Haskel, Pereira and Slaughter [48] for UK concluded that spillovers depend on productivity gap. Barrios et al., [49] for Spain, Yudaeva et al. [50] for Russia, Marin and Bell [51] and Chudnovsky, López and Rossi [52] for Argentine provided for the important role of absorption capacity in the manifestation of spillovers. Aitken and Harrison [13] for Venezuela, Girma and

⁴ In general, the studies focused in transition countries of Central and East Europe examine spillovers in the manufacturing sector, with the exception of Djankov and Hoeckman (2000) and Konings (2000), who include also non-manufacturing sectors.

⁵ One possible explanation for this is that researches focused on transition economies use relatively small databases and/or short time periods, given the short experience of FDI in these countries.

Wakelin [53] for United Kingdom, and Ivarsson (2002) for Sweden, provided evidence for geographical proximity as an important factor for the generation of spillovers suggesting that it constituted in increasing business linkages between foreign firms and domestic ones. The results provided by empirical literature in the third group confirm the statement of Blömstrom [55] “*spillovers depend crucially on the conditions for local firms*” (p: 177, [55]).

The fourth group of studies includes the examination of inter-industry spillovers, which represents quite a new area of analysis, with papers written only during the recent years. Despite the intense interest of policy makers in the subject, Blomström, Kokko and Zejan [56] postulate that there are hardly any empirical papers working explicitly on the existence of vertical spillovers, which might be a forward or backward relationship. It is generally noted that vertical spillovers may be more important than horizontal ones, however, due to data limitations, statistical analysis on these effects was lacking. Vertical spillovers were generally examined through case studies. Notable exceptions are the recent studies of Kugler [57] for Columbia, Blalock [51] for Indonesia, Batra et al. [51] for Malaysia, Chudnovsky, López, and Rossi [52] for Argentina, Schoor and Van der Tol [60] for Hungary, Smarzyska [22] for Lithuania, Smarzyska Javorick et al. [61] for Romania and Yudaeva et al., [50] for Russia. Most of these researchers, except that of Yudaeva et al., [50], claim that inter-industry spillovers’ mechanisms are successful and are likely to operate more efficiently than intra-industry spillovers.

It is evident that there is huge controversy in the empirical results and there is little conclusive evidence to support the beneficial effects of FDI to the host companies. Pioneer studies on productivity spillovers provided evidence of positive intra-spillovers effects, using cross-sectional industry level data. However, recent empirical research on spillovers strongly suggest for mixed results, applying not only for developed countries, but also for developing and transition economies. Some of them revealed negative productivity spillovers, while others showed that spillovers may exist but they are highly dependent on various conditions of domestic firms. Empirical literature on the subject has emphasized the importance of using panel data as the correct way to detect for spillovers. Scholars have offered various explanations on mixed results. Some of the most convincing ones are that different studies have applied different methodologies, have focused on different countries, and have studied effects of different MNEs. Spillovers are a difficult concept in itself and this has made the research even more complicated.

3.2 Review of Survey Approach – Case Studies and Sample Surveys

Apart from econometric studies, a number of qualitative surveys have been conducted in various countries so as to examine the direct and indirect technological effects of FDI. Surveys dealing with direct and indirect technological effects of FDI include case studies and sample surveys (questionnaires and interviews). Case study approach includes comprehensive and rich descriptions about particular FDI companies or projects in specific countries, providing firm specific information on the role of MNEs in changing the productivity of local firms. Case studies were particularly important during the seminal work on FDI spillovers, putting emphasis to linkages, labour turnover and demonstration effects [57]. As Lipsey and Sjöholm [62] put it, “case studies offer great flexibility” (p: 31, [62]). This is extremely important for the analysis of the specific subject as the nature of technology transfer differs across companies, industries and countries. On the other hand, sample surveys provide detailed information on FDI effects by reporting results from in-depth questionnaires and/or interviews with companies’ chief officers (CEO) or managers. Most importantly, surveys provide information on the mechanisms through which FDI effects diffuse in the host country, which is a gap in the econometric studies. Many researchers favour this method also due to lack of comprehensive firm level data or even incomplete records of financial data, which is the case particularly for transition countries. However, in spite of detailed qualitative information provided by these approaches, they have a number of drawbacks. Generalizations with respect to the results of sample studies and particularly of case studies are not easy. On the other hand, surveys involve a degree of subjectivity; hence their results should be approached with caution.

3.2.1 Case Studies

A number of researches include case studies on single or a number of companies and/or sectors. These provide important information on the subject; however, case studies are difficult to be generalized. Lall [63] was among the first who found case study evidence to support vertical linkages between local firms and MNEs in the study of truck manufacturers in India. The effects were particularly strong towards upstream local suppliers, rather than downstream customers. The author argued that foreign affiliates provide support to domestic suppliers in various ways: by training employees and managers, by helping them in upgrading production facilities, by providing technical assistance in order to improve product quality, and by providing support in purchasing raw materials. Following Lall [63], a number of case studies emerged covering developed, developing and transition countries. Kugler [57], Rhee and Belot [64], Moran [65], UNCTAD [66] and Haskel, Pereira and Slaughter [48] have provided a summary of case studies on developed and developing countries concluding that case studies at firm and industry level suggest that spillover effects from MNCs vary according to technological and managerial capabilities of local firms relative to those of foreign firms, as well as to strategies followed by the latter.

Most of the recent case studies combine detailed sample surveys with some case studies in order to provide more generalizable results. For instance, Altenburg [67] presented three case studies from manufacturing activities (electronics hardware, automotive and apparel industries) in developing Asian countries. Even though, many suppliers were eliminated or crowded-out due to their low quality and inefficiency, especially in the manufacturing industry, Altenburg suggested that as spatial proximity to suppliers and customers becomes an important advantage, MNEs are interested in intensifying linkages with neighboring firms. As a result, new opportunities were coming up for specialized local partners and suppliers. Parallel to this, competition pressures, forced MNEs to put higher requirements to their suppliers leading to an improvement in quality, price and time delivery. Pavlenik and Smith [68] used a case study approach at industrial level for the Czech and Slovak Republic and found some embeddedness but which was created through transformation of existing systems and not from vacuum. Deeper forms of embedded institutional reforms, as well as enrichment of skills, increasing wages and productivity and high levels of cooperation and partnership were largely absent. Using the same approach, Pavlenik [69] found for Czech automobile industry that FDI had only limited impact on domestic R&D and that the rumours of large-scale technology transfer of industrial R&D by MNEs were exaggerated.

3.2.2 Sample Surveys

A bunch of studies have referred in details, by conducting in-depth surveys, the direct and effects of foreign companies. Regarding the direct effects, survey evidence has shown that most of the FDI projects in developed, developing and CEE region result in technology, knowledge and skills transfer, as well as improvements in quality and productivity. As a result, it is not surprising that foreign affiliates perform better than domestic firms (particularly in the manufacturing sectors which is extensively analysed). Moreover, regarding especially the CEE region, FDI either through M&As, Greenfield investments or privatisations, has played important role in the restructuring and upgrading of domestic companies, mostly through increase in quality and productivity. Studies supporting such findings for developing countries are those of Mirza, Giroud and Köster [70] and Mirza and Giroud [71] for the Asian countries. Whereas for transition countries, Zemplerova [72] for Czech Republic, Hunya [73] for Poland and Elteto [74] for Hungary confirmed the better performance of foreign firms compared to that of domestic firms. Moreover, Zemplerová and Jarolim [75] for the Czech Republic found that FDI played an important role in enterprise restructuring while Rojec ([76], [77]) in a comparative study for the Czech Republic, Hungary, Poland and Slovenia found evidence that FDI transferred new technology, know-how and finance, as well as supported the company's access to foreign markets.

Concerning the indirect effects, there is no conclusive evidence on technology spillovers from foreign firms to the domestic firms, with some studies providing evidence on positive spillovers (particularly on backward linkages with suppliers) and some others failing to show any benefit of host companies from this phenomenon due to the fact that the materialization of these effects need time to emerge,

combined with the fact that they depend strongly on host country characteristics, host company characteristics, as well as the characteristics of foreign enterprises investing in the economy. Some of the studies that have showed positive FDI indirect effects are those of PA Cambridge Economic Consultants [78] and Potter et al., [79] for UK, Crone and Roper [80] for Northern Ireland, Blalock and Gertler [26] for Indonesia, FIAS [81] for Latvia and Ferencikova [82] for Slovakia. Mirza, Giroud and Köster [70] found mixed results for five ASEAN countries. Finally, one of the most recent studies is carried out by Smarzynska Javorcik and Spatareanu [61], which controlled for spillover effects to domestic competitors through interviews with domestic firms in Latvia and the Czech Republic. The study confirmed the existence of spillovers through knowledge transfer (the size and type of effects were slightly different in the two countries). The main mechanisms channeling this transmission were the demonstration effects and the movement of labor. Competition effect was also important, even though adverse effects were found on some firms in the short run.

4. Developments of FDI on the Albanian Economy

A transition country like Albania, with small size and strategic location, constitutes a very interesting case for investigating the direct and indirect impact of FDI. Anyone can imagine that Albania is too far away from being a technological advanced economy and needs technology to stimulate competitiveness, industrialization and development. Indeed, being one of the poorest countries of South-East Europe and struggling to improve its economic condition, FDI could bring technology and could play a critical role in the upgrading of local companies and in the development of the country. The Albanian government and politicians feel optimistic about the benefits that the country can enjoy from FDI, which has already been attributed an important role in the transformation of the transition economies, combined with high expectations of local governments [21], [8]. Adding to this, the availability of low cost and well-educated labour force (inherited from the communist regime) and the proximity of Albania with the developed European economies make it a place particularly likely for FDI effects to manifest themselves. Despite this, it is interesting that in contrast to other CEE countries, the impact that FDI can have in Albania has not attracted attention of researchers and is a missing component in the voluminous FDI literature. Albania lacks evidence on direct and indirect technological transfer of FDI to host companies and little is known on the role that foreign companies can play on the country, which gives a strong incentive to examine the subject in this specific country. Given its future potential for FDI growth, Albania constitutes a real challenge for researchers. Evidence from other countries of the region might also appear helpful in explaining and understanding the role of FDI and the way that technology diffuses in the host economy.

Nevertheless, before examining the FDI impact in Albania, it is important to provide information on developments of FDI in the country. Since the beginning of its restructuring, the FDI attracted by the Albanian economy has been increasing through the years, particularly after 1998. However FDI in Albania stays far below the volume experienced by other Central and Eastern European Countries. FDI in Albania remains limited mainly due to a relatively insecure investment environment, poor infrastructure, heavy administrative procedures, corruption in the public administration and the judiciary and relatively high taxes. Nevertheless, there are potentials for FDI growth in Albania due to countries main competitive advantages such as favourable natural conditions, proximity to key EU markets, relatively low cost but skilled labour force and progress in the privatisation process.

In 1995 and 1996, FDI annual inflows to Albania were nearly twice as large as the period from 1997 to 1999. This decline in FDI was certainly the consequence of a series of crisis that affected the country, starting with the 1997 civil disturbances that followed the collapse of the pyramid schemes, the coup attempt in September 1998, and the Kosovo crisis in 1999. The level of FDI during 1998-1999 amounted only 1% of GDP. However, its level has increased substantially since then; in 2000 the volume of FDI has been estimated to be around USD 143 million, which is three times higher than in 1999. In 2000, FDI activity increased not only in terms of volume but also in terms of number of foreign companies investing in Albania. The number of registered joint ventures and foreign wholly owned firms increased by about one-third, between 1999-2000. Most of the increase in FDI activity came as a result of privatization programs undertaken in the telecommunication and banking sectors.

The FDI level continued to increase for 2001 and can be estimated around USD 207.0 million. This progress was notably the result of successful privatizations, such as: AMC (Albanian Mobile Telecommunication), Vodafone and Albkrom (Chromium Albanian Industry), and Albcooper (Cooper Albanian Industry privatized by the Italian company DARFO). For the year 2002 the FDI estimated has been USD 150.0 million. Compared to that of 2001, the FDI figure has decreased as a result of the delays in privatisation deals for strategic objects such as Albtelecom, Savings Bank, etc. For 2003, FDI amounted USD 178.2 million, where the privatisation of the Savings Bank by the Austrian Raiffeisen Bank had the major contribution. The FDI stock for the period 1992-2003 has reached USD 1.13 billion. Estimated figures for 2004 are around USD 300 million. Nevertheless, Albania continues to attract less foreign investment in comparison with other countries in the region. By the end of 2003, FDI in Albania represented only 0.86% of the total FDI in Central and East Europe.

Table 1 illustrates FDI inflows in Albania for the period 1992-2003 based on data provided by the Bank of Albania. Other organizations, such as EBRD and IMF, provide slightly different statistical data on FDI inflows, implying for the difficulties faced in the exact measurement of FDI inflows. Nevertheless, the data that are presented in this section for FDI flows, sectors of investment, location etc. are based on national sources, such as National Institute of Statistics (INSTAT) [84] and Bank of Albania [85].

Year	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Total	32.0	45.0	53.0	70.0	90.1	47.5	45.0	41.2	143.0	207.0	150.0	178.2	300.0

Table 1 FDI Inflows in Albania (in USD million). Source: Bank of Albania (2004).

FDI as a percentage of GDP has increased since 1999, showing an upward trend until 2000 accounting for 5% of GDP and then falling in 2002 up to 3.2% (see Figure 2).

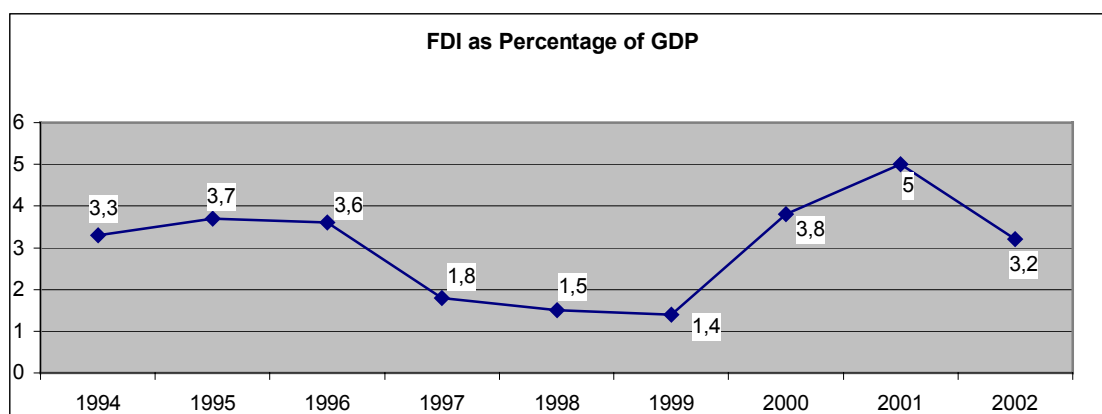


Figure 2 FDI as a Percentage of GDP in Albania. Source: Author's Calculations.

For the years to come, the FDI level will continue to depend on the privatisation process due to the fact that the statistical data of the previous years demonstrate that privatisation is the major source of foreign investments in Albania. When the privatisation process will finish, major efforts will be dedicated in attracting FDI with 100% foreign capital and in promoting form of cooperation consisting in foreign capital as well as domestic one.

In countries of CEE, predominant are global investors. In contrast to other CEE countries, FDI in Albania originates mainly from neighbouring countries, such as Italy and Greece. The main origin countries for FDI for the year 2003 were Italy (47.9%), Greece (34.2%), FYROM (2.2%) and Turkey (2%) (see Table 2). The remaining is from other European and US countries. Albania receives the main part of FDI (87%) from the European Union, which is even the major trade partner for the country.

The Italian investments are mostly located in West of Albania, close to the Adriatic Sea. The majority of the Italian investments are small and medium enterprises (SMEs), engaged mainly in construction

(35%), textile and shoes production (21%), trade and services (16%), and agricultural food processing industry (8%). Italian companies take advantage and make profits from the Albania's comparative advantage in low cost quality labour force; Albanian wages are ten times lower than those in Italy. Moreover, Italian investors take advantage of the Italian Government grants and subsidies designed to promote Italian investments in Albania. According to the statistical data provided by the Bank of Albania (2002), there are approximately 500 Italian companies operating in Albania in forms of joint ventures and wholly owned companies. The private Italian investments are estimated to have a value of above USD 400 million. The most important Italian companies are: DARFO in the area of chrome industry: ENEL, ESSEGEJ, BEGHETTI in the area of Hydropower.

Source Countries	%	Source Regions	%
Italy	47.9	European Union (15)	87
Greece	34.2	USA	2
Germany	1.3	Far East	2
France	1.3	Middle East	2
United Kingdom	1.1	Eastern Europe	1
Belgium	0.9	Regional Countries	6
Austria	0.2		
US	2		
China	1.3		
Turkey	2		
FYROM	2.2		
Kosovo	1.1		
Croatia	0.7		
Bulgaria	0.2		
Spain	0.2		

Table 2 FDI in Albania According to the Source Countries and Regions (2003) (as a % of total). Source: Bank of Albania (2004).

Greek investments are mostly concentrated in Tirana and in the south and southeast of Albania, close to the Greek border. In general, Greek investors in Albania are concentrated on trade and recently on telecommunications, banking and construction, and only less than 2 % are involved in other branches of industry such as textiles, garments, manufacturing of leather products and tobacco processing. More than 60 % of Greek foreign investors benefit from grants offered by the Greek governments to the Greek companies operating in the Albanian market. Actually, according to the Bank of Albania, there are around 213 Greek companies, making 34.2% of the foreign investments in Albania. The major part of the Greek investment is concentrated in cities such as: Tirana, Korca, Gjirokastra, Devoll and Fier.

Foreign investments are mainly concentrated in the main districts of Albania, which are Tirana (55 %) and Durres (12%) [84]. These are the two largest cities in the country and account for approximately 67 % of the total companies operating with foreign or joint capital. Durres is responsible for the largest port handling of the import-export activities. On the other hand, Tirana as the capital city of Albania is more exposed to investor and is becoming the country's business centre. Other Albanian cities that have attracted FDI are Korca (9%), Gjirokastra (6%), Shkodra (4%), and Vlora (3 %).

Figure 3 shows the distribution of FDI according to sectors. Besides the fact that there are a large number of foreign enterprises involved in trade, there are only few amounts of FDI inflows in this sector. FDI in agriculture remains low due to small and fractured land plots, restrictions on foreign ownership, of rural land, and weak roads and transport links to major national and regional markets. Sectors that attract more foreign investments in agriculture and related industries include fisheries, fish processing and canning, food processing, olive oil refining, beverages production, and wood processing. Regarding services, FDI is concentrated in banking, retail and construction. In the construction sector, FDI has grown in terms of the number of enterprises and volume of activity. The origin countries of most of the foreign investors in construction are from Turkey, Italy, and Greece. Based on the flow of resources from donor countries and in response of the significant rural to urban

migration, there are prospects for FDI to expand in this area, as there is growth in investments in infrastructure and buildings. Finally, the majority of foreign enterprises operating in manufacturing sector are involved in re-exporting finished and semi-finished goods, such as garments, small leather articles, and shoes. Re-exporting takes place when foreign affiliates in Albania import raw materials from abroad (usually the investor's home country), process them using labour intensive production methods that capitalize in Albania's low labour cost, and after that re-export the goods at various stages of processing (often back to the investor's country of origin).

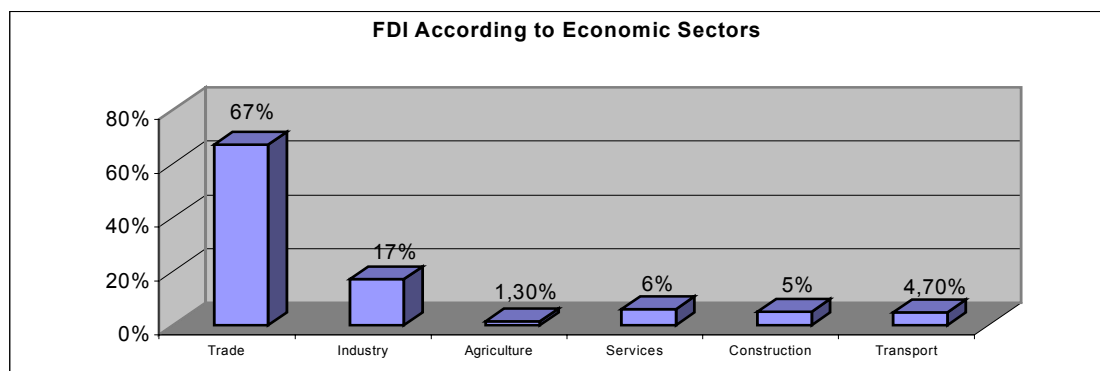


Figure 3 FDI in Albania According to Economic Sectors (2001). Source: INSTAT (2002).

To sum up, Albania was the only country in the Balkans that experienced communism in its more severe form, being highly isolated from the rest of the world throughout the regime. However, escaping from the communist regime, Albania has gone through a period of transformation from a centralized to a market economy, since the early 1990s. Despite its delay in the economic development, an on-going privatisation process and a significant potential for increasing FDI inflows characterize the country. Even though FDI levels in Albania remain low compared to other countries of the region, FDI already constitutes an important component of the private sector development, and consequently a critical component of the Albania's economic development. The Albanian government has created a legal framework with incentives for foreign investors, and has undertaken a number of measures to support foreign investors. As a result of this, combined with macro-economic and politic stabilization, the FDI inflows into the country have increased sharply since 1998, with major investments from Italy and Greece. Given the increasing presence of foreign companies in Albania, a number of interesting questions arise: What is the role of these companies in Albania? Does FDI serve as a channel of technology transfer to the local companies? Does FDI support the upgrading of the local companies?

5. Methodology

Despite the fact that the research on FDI technological effects began almost three decades ago, the two main approaches used so far to provide empirical evidence on the subject are the survey approach and the econometric approach. In the absence of firm-level data needed for econometric analysis, this research employs an explanatory study utilizing sample survey study. According to certain scholars, surveys are superior to econometric studies as they provides comprehensive description and rich data on FDI effects and on the mechanisms through which these effects diffuse in the host country (Van de Ven [86], offering answers to the questions that econometric analysis fails to provide. The analysis is based on data from two surveys carried out during 2004 and 2005 by the Albanian Government in cooperation with the Albanian Institute of Statistics. The sample surveys serve three purposes: 1) To provide evidence on direct effects and indications for indirect effects of FDI; 2) To answer to the questions how and why technological transfer takes place through FDI; and 3) To construct different typologies ranging in the scale and type of impact. Thus, the major contribution of the paper is that, through the survey research, it will attempt to hopefully extend knowledge and to offer a better approximation of the issue, by providing extensive and accurate information in order to answer the research questions and explain why direct and indirect technological transfer has been more effective

in some industries than others, as well as to assess whether these successes have transferable lessons for other industries. In implementing this methodology, the study will be in line with the studies of PA Cambridge Economic Consultants [78], Crone and Roper [80], Mirza, Giroud and Köster [70], Mirza and Giroud [71], Ferencikova [82] and Smarzynska Javorcik and Spatareanu [61].

The explanatory study includes detailed surveys with foreign companies in the Albanian manufacturing sector including all sizes, entry modes, nationalities. Firstly, pilot examples with 6 foreign companies were carried out during spring of 2005 with the purpose of finalizing the questionnaires. Secondly, the sample survey was carried out during January-April 2006 covering all active foreign companies operating in the four cities that have attracted more FDI in Albania which are Tirana (64,6 %), Durrës (11,8), Korça (7,5%), Elbasan (2,5%) (based on data from the survey conducted by INSTAT and Bank of Albania during 2005). The questionnaires were addressed to approximately 190 foreign companies. Eighty-seven of them agreed to participate in the survey (providing a response rate of 46 %), while the rest either refused to participate, or couldn't be found due to change of addresses, or were shut down due to financial problems. Detailed face-to-face structured interviews were carried out with the general directors or production managers of foreign companies covering the following lines: the activity of foreign companies and their characteristics; the extent to which foreign companies bring with them technology, the impact of foreign companies on the practices and performance of local suppliers, customers and competitors; the mechanisms through which technology is transferred to local companies; the mechanisms through which technology is transferred to local companies; FDI and non-market collaborations; and the wider effects of FDI on local economy. As already mentioned, surveys were conducted through structured interviews or standardized interviews or interviewer-administered questionnaires, which were selected as the best method to approach this study as compared to other types of questionnaires since they provide higher quality and reliability of the data [87]. They usually have a higher response rate than self-administered questionnaires, and enable the interviewer to ensure that the respondent is the person wanted to answer the questionnaire and to avoid any misunderstanding and confusion with reference to questions (the interviewer can clarify on the spot ambiguous questions). Moreover, establishing personal contact is of high importance.

Thirdly, in order to provide even more qualitative insight on FDI effects, particularly on indirect or spillover effects, survey studies were followed by case studies, which are used as a second method for the study. Case studies were selected according to the relative importance of the sectors. The three most sectors were beverages industry, furniture industry and production of construction materials. Two foreign companies were selected for each sector, and the former covered different typologies. Case studies include detailed open-questioned interviews with foreign companies, complemented with structured questionnaires with their local competitors, suppliers and customers. Structured questionnaires with local companies were completed out either through face-to-face interviews or interviews through the phone. Case study approach is considered to be a very good method in identifying the upgrading of the domestic companies through FDI technological transfer and particularly in answering the explanatory questions of how and why [88]. This is because case study analysis, through the use of multiple sources of evidence (allowing for quantitative and qualitative analysis), is thought to explain a contemporary phenomenon within its real-life context when there are no clear-cut boundaries between the phenomenon and the context and when the researcher has little control over events [p: 123, [88]]. "Case studies offer great flexibility" (p: 31, [62]), which is of high importance for the analysis of our specific issue as the nature of technology transfer differs across companies, industries and countries. In contrast to surveys that are based on statistical generalization, case studies are based on analytical generalization, which implies that results from case study approach should be treated with caution given that often the sample cases chosen for the analysis are not representative [88].

6. Preliminary Results⁶

6.1 Characteristics of the Survey Sample

The questionnaires were addressed to approximately 190 foreign companies including all active manufacturing foreign companies operating in the four cities (Tirana, Durres, Korca, Elbasan) that have attracted more FDI in Albania. Eighty-seven of them participated in the survey, providing a response rate of 46% (see Table 4 for details on the companies included in the sample). More than 26% of the companies included in the sample operate in the textiles and clothes industry, followed by 16% in the food industry, 9% in the manufacturing of electrical materials, 8% in the production of construction materials, 8 % in the manufacturing of electrical materials, 8% in the paper industry, 6% in the shoes industry and the 10% in other industries. More than 41 of the foreign companies have more than 100 employees, which is characteristics mainly for the textiles and shoes industry. The foreign investors originated from nine different countries. Italy was the predominant origin country followed by Greece. 55.2% of survey companies originated from Italy, 31% from Greece and the rest 13.8% from Germany, Austria, France, Bulgaria, Serbia, USA and Liban. More than half of foreign companies were established in Albania before 2000, which is particularly important given that academics support the idea that FDI indirect effects need time to manifest themselves. Greenfield investment turned out to be the most popular mode of investment in Albania accounting for 62.1%, followed by joint ventures 23% and then by acquisitions 10.3%. 70% of the foreign companies are owned wholly by foreign investors, while the rest are joint ventures. More than half of the companies included in the sample are located in Tirana, 18% in Durres, 18% in Korca and 7% in Elbasan.

Characteristics	No.	%
Main activity (sector of operation)		
Textiles and clothes industry	23	26.4
Shoes industry	5	5.7
Wood industry and manufacturing of furniture	7	8.0
Food industry (food, beverages, and tobacco)	14	16.1
Paper industry (production of paper, publishing and printing houses)	7	8.1
Manufacturing of electrical materials	8	9.2
Production of construction materials	7	8.1
Manufacturing of metals	7	8.1
Others	9	10.3
Total	87	100
Employment size		
10-50	27	31.0
50-100	24	27.6
Above 100	36	41.4
Total	87	100
Origin country		
Italy	48	55.2
Greece	27	31.0
Germany	3	3.4
Austria	3	3.4

⁶ The results presented in this section should by no means be considered as definitive as the research is currently in progress and the results are only preliminary.

France	2	2.2
Bulgaria	1	1.2
Serbia	1	1.2
USA	1	1.2
Liban	1	1.2
Total	87	100
Year of establishment in Albania		
Before 2000	45	51.7
2000-2004	42	48.3
Total	87	100
Mode of establishment in Albania		
Greenfield	59	62.1
Acquisition	22	10.3
Other	6	4.6
Total	87	100
Ownership of subsidiary		
Joint-venture	26	30.0
Wholly-owned	61	70.0
Total	87	100
City in Albania		
Tirana	44	50.5
Durres	18	20.7
Elbasan	7	8.1
Korca	18	20.7
Total	87	100

Table 4 Characteristics of the survey sample

6.2 Direct Effects of FDI - Human Resource and Technology

Regarding the direct FDI effects on human resource and technology, in general, structured questionnaires showed that these effects have been significant particularly in terms of employment, training, reinvestment, and technology. It is important to note that 62% of foreign companies had a Mother company in the home country, while the rest were “stand alone” companies. In addition, only 5 % of foreign companies had a sister company operating in Albania.

More than 41% of foreign companies are large employing more than 100 employees, which results in a direct impact of FDI to local employment. A very high proportion of workers, particularly in the textiles industry, are women.

Foreign companies provide different ways of training to their employees. First, the most common method of training is on-the-job-training offered to employees through learning by doing/performing and learning by demonstration. Regarding the newcomers they have to go through a training period of more than 2 months, depending on the company and on the type of employee. Second, external staff sent by the Mother Company offers training to employees particularly when the subsidiary company acquired new machineries or new designs. Most of the foreign companies benefit from site visits of external staff, including 2 or 3 experts per year for a period of 1-2 weeks, depending on the needs of the company. Third, some of the companies offer external training to employees by participating in training seminars organized by training institutions and business associations. Fourth, training is

offered externally outside the country, in most of the cases in the origin country of the foreign company. The third and fourth way of training involved staff holding the position of director, manager and technician. However, it is interesting to remark that only 2% of the companies included in the survey have a separate training department.

The foreign companies were satisfied from the level of technicians and simple workers, however, highlighted the lack of skilled and educated managers. As expected, they confirmed a high movement of workers from foreign companies to local companies and vice-versa implying for indirect spillover effects.

Concerning technology, the survey shows that all foreign companies have bought their production machinery from abroad, in some cases purchased new and in some others second hand. It was interesting to note that although many foreign companies started their activity with machinery second hand, due to increasing competitive pressures and increasing expectation of the local market, they decided to introduce new ones. In general, Mother Company and foreign partners (if any) transferred new technology (in the form of machinery and equipment), capital, skills and know-how (in the form of expertise and training), and promoted exporting of foreign subsidiaries (exporting of products either to mother company either provide links with customers outside the host country).

More than 60% of the foreign companies claimed that they were reinvesting their profits, however, only few companies were able to give precise figures.

6.3 Indirect Effects of FDI to Local Suppliers, Customers and Competitors

Technology transfer to local companies is limited mainly due to little contact of FDI companies with the host economy. 45% of the foreign companies included in the survey are subcontracting companies involved in manufacturing sectors such as textiles and clothes, shoes, and electrical equipment industry. Textiles and shoe making industries are a tradition in the Albanian industry since the 1980s. The majority of these companies are involved in re-exporting finished and semi-finished goods usually back to the investor's home country. These companies use labour intensive production methods, taking advantage of Albania's low labour cost, to process the goods that are re-exported at various stages of processing. It is evident that these foreign companies operate in enclave economies without having any contact with local suppliers, customers and competitors. The only competition with local firms is in attracting and keeping local workers in the company due to high movements of workers amongst competitive firms. However, even in these cases, there is local value-added (wages and salaries) and large-scale of production and employment (most of textile and shoes companies employ large number of workers).

55% of the foreign companies included in the sample are not subcontracting companies and 40% of these companies purchase at least 10% of their inputs locally and 56% of these companies sell at least 10% of their products to local customers (many of these companies are export-oriented and intend to extend their exports even more in the future). Those few foreign companies that source their inputs or part of their inputs locally ensure to provide support to their local suppliers by assisting particularly in terms of quality, price, speed of delivery, and method of distribution. This results in improvements in the performance of local suppliers through increase in sales and productivity. Local customers that purchase the products from foreign companies also benefit, however, at lower levels than local suppliers. Local customers benefit directly by improving their purchasing practices and most important by purchasing better products from foreign companies, which results in improvement in quality, reduction in price and increase in sales. Regarding local competition, more than 60 % of the non-subcontracting companies declared to face significant direct local competition highlighting the high competition in prices from the black market. Foreign companies claimed to have competitive advantages (quality, technology, design, expertise, reputation) compared to legal local competitors. A number of them confirmed that local competition has been increasing significantly in the recent years; local competitors bring technology and know-how from neighbour countries such as Italy and Greece. This puts pressures for reduction in prices, as well as increase in the variety of products. The competition level is expected to increase even more in the coming years. Most of the foreign

companies provided evidence for demonstration effects, claiming that local companies try to imitate their technology and products. The demonstration effects have been more successful in terms of product design; on the other hand, it is difficult to imitate quality due to high technology gap between foreign and domestic companies. In general there are few links of foreign companies with local suppliers, customers and competitors, however in case when they exist, indirect effects are more evident in wood industry, food industry and production of construction materials.

Main transmission mechanisms for spillover effects of foreign companies on local companies involve direct cooperation of foreign companies with local ones (particularly suppliers and customers) mainly through active mechanisms such as site visits on technical and quality issues, informal sharing of views and ideas, higher requirements put by foreign companies, as well as through passive mechanisms such as demonstration/imitation effects (particularly competitors).

6.4 Infrastructure and Non-Market Collaborations

All foreign companies reported that they are highly dissatisfied from the Albanian infrastructure, pointing out the poor level of energy supply and transport services. Moreover, they expressed their concern on the poor role that Albanian government has played so far on supporting the activity of FDI, where there is lack of financial incentives and tax relief.

Many foreign companies participate at training seminars and programs organized particularly by the Albanian Chamber of Commerce, the American Chamber of Commerce and Foreign Investors in Albania Association (FIAA) and are involved in various sponsorships and charities. Some of these companies are also benefiting from U.S. Agency for International Development (USAID), which provides technical assistance and loans to small and medium companies in Albania. However, links of foreign companies with local universities appear to be limited.

7. Conclusions

Many countries offer incentives and are actively encouraged to attract FDI into different sectors of economy, based on the belief that FDI improves technological standards along with efficiency and competitiveness of the host economy; brings in the latest technology; and promotes productivity growth by bringing in the host country a bunch of productive assets such as long-term foreign capital, technology, skills, entrepreneurship, innovative capacity, as well as organizational, managerial, marketing and export know-how. This takes place throughout a process of technological transfer from FDI to the companies of the host country in two ways: directly and indirectly. Foreign enterprises constitute in a direct injection of foreign capital, technology and foreign management skills to their affiliates, which in turn leads to higher productivity. This is referred as the direct effect of FDI. On the other hand, the indirect effects that are referred as spillovers or externalities, can occur between FDI companies and their local competitors (intra-industry spillovers) or between FDI companies and their local suppliers (inter-industry backward spillovers) or between FDI companies and their local customers (inter-industry forward spillovers). Literature has found mixed empirical evidence on the existence of spillovers. Evidence is strongest in case of backward linkage with local suppliers (foreign firms provide them with technical and financial assistance, training and other support).

Albania is the only country in the Balkans that experienced communism in its more severe form, being highly isolated from the rest of the world throughout the regime. In its efforts to improve its economic and technology standards, Albania can serve as in vitro case study to trace back straightforwardly the technological effects of FDI, given its small labour-intensive economy, low cost but skilled labour force and the fact that the major investors in Albania originate from neighbouring countries. Preliminary results from sample survey with foreign companies in the manufacturing industry indicate that a considerable number of FDI companies in Albania, such as subcontracting companies in textile and shoes industry, are involved in low-value added activities, utilizing labour-intensive production. These activities are associated with minimal investment in technology and capital. Most of these foreign companies avoid majority investments and involvement in the Albanian economy, and the

linkages with other Albanian companies or institutions are barely existent. However, this is not the case for the non-subcontracting companies (particularly in the wood industry, food industry and production of construction materials), which have brought new technology, know-how and skills in Albania, implying for important direct effect of FDI. Nevertheless, the technology transfer to other local companies is limited mainly due to little contact of FDI companies with the host economy. In case where foreign companies have contacts with the local economy, local suppliers benefit (assistance on quality, price and time delivery) more than local customers and local competitors. Positive effects to local competitors are balanced by adverse effects (crowd-out effects due to increasing competitive pressures). Main transmission mechanisms for the indirect effects involve direct cooperation mainly through informal sharing of views, higher requirements put by foreign companies, as well as through demonstration/imitation effects.

The preliminary results of the study emphasize the necessity for the Albanian government to create a right mix of policy measures and incentives to attract as well as keep in the country serious and efficient foreign investment, associated with conditions favourable for the emergence of FDI effects. This will in turn, improve the image of Albania as a potential host country by promoting its existing competitive advantages and also developing new ones.

References

- 1 Hunya, G. Recent Impact of Foreign Direct Investments on Growth and Restructuring in Central European Transition Countries. Vienna Institute for International Economics, (2002).
- 2 IMF. Balance of Payments Manual. 1993, Washington D.C.: IMF
- 3 OECD. Benchmark Definition of Foreign Direct Investment. 1996, Paris: OECD Publications.
- 4 Hymer, S. The International Operations of National Firms: A Study of Foreign Direct Investment. (1976), Cambridge: MIT Press.
- 5 Dunning, J. H. International Production and the Multinational Enterprises. 1981, London: Allen and Unwin.
- 6 UNCTAD World Investment Report. FDI Policies for Development: National and International Perspectives. 2003, New York and Geneva.
- 7 Blomström, M. and Kokko A. Multinational Corporations and Spillovers. Journal of Economic Surveys 1998; 12: 1-31.
- 8 UNECE. Economic Growth and Foreign Direct Investment in the Transition Economies. 2001, UNECE Economic Survey of Europe.
- 9 Hoekman, B., Maskus K. and Saggi K. Transfer of Technology to Developing Countries: Unilateral and Multilateral Policy Options. 2004, World Bank Policy Research Working Paper.
- 10 Portelli, B. and Narula R. Foreign Direct Investment Through Acquisitions and Implications for Technological Upgrading. Case Evidence from Tanzania. 2003, Globalization as Transformative Process Working Paper Series.
- 11 Blomström, M. and Kokko A. FDI and Human Capital: A Research Agenda. 2002, OECD Technical Papers.
- 12 Harris, R. and Robinson C. Foreign Ownership and Productivity in the United Kingdom: Estimates for UK Manufacturing Using the ARD. Review of Industrial Organization 2003; 22: 207-223.
- 13 Aitken, B. and Harrison A. Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela. American Economic Review, 1999; 89: 605-618.
- 14 Rotenberg, J. and Saloner G. Competition and Human Capital Accumulation: A Theory of Interregional Specialization and Trade. Regional Science and Urban Economics, 2000; 30: 373-404.
- 15 Mirza, H. and Giroud A. Regional Integration and Benefits from Foreign Direct Investment in ASEAN Economies: The Case of Vietnam. Asian Development Review, 2004; 21: 66-98.
- 16 Teece, D. J. Technology Transfer by Multinational Firms: The Resource Cost of Transferring Technological Know-How. Economic Journal, 1997; 87: 242-261.
- 17 Görg, H. and Greenway D. Much Ado About Nothing? Do Domestic Firms Really Benefit from Foreign Direct Investment? 2002, University of Nottingham Working Paper.

- 18 UNCTAD World Investment Report. Transnational Corporations and Export Competitiveness. 2002, New York and Geneva.
- 19 Schoors, K. and Van der Tol B. Foreign Direct Investment Spillovers Within and Between Sector: Evidence from Hungarian Data. 2002, Ghent University Working Paper.
- 20 Dunning, J. H. Multinational Enterprises and the Global Economy. 1993, Wokingham: Addison Wesley.
- 21 UNCTAD Promoting Linkages Between Foreign Affiliates and Domestic Firms. 2001, Geneva: United Nations Conference on Trade and Development.
- 22 Smarzynska, B. Does Foreign Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages. 2002, World Bank Policy Research Working Paper,.
- 23 OECD New Horizons and Policy Challenges for Foreign Direct Investment in the 21st Century. 2001, OECD Paper.
- 24 Meyer, K. FDI Spillovers in Emerging Markets: A Literature Review and New Perspectives. 2003, DRC Working Paper.
- 25 Kugler, M. The Diffusion of Externalities from Foreign Direct Investment: Theory Ahead of Measurement. 2000, University of Southampton Working Paper.
- 26 Blalock, G. and Gertler P. Welfare Gains from Foreign Direct Investment through Technology Transfer to Local Suppliers. 2004, University of California Working Paper.
- 27 Damijan, J. P., Majcen, B., Knell M. and Rojec M. Technology Transfer Through FDI in Top-10 Transition Countries: How Important Are Direct Effects, Horizontal and Vertical Spillovers? 2003, WDI Working Paper.
- 28 Haddad, M. and Harrison A. Are There Positive Spillovers from Direct Foreign Investment? Evidence from Panel Data for Morocco. *Journal of Development Economics*, 1993; 42: 51-74
- 29 Barrell, R. and Pain N. The Growth of Foreign Direct Investment in Europe in R. Barrell and N. Pain (eds.), *Innovation, Investment and the Diffusion of Technology in Europe*, German Direct Investment and Economic Growth in Postwar Europe. 1998, Cambridge: Cambridge University Press, 19-43.
- 30 Plaffermayr, M. Foreign Outward Direct Investment and Exports: A Time Series Approach. *Applied Economics*, 1994; 26: 337-351.
- 31 Bellak, C. How Performance Gaps Between Domestic Firms and Foreign Affiliates Matter for Economic Policy. *Transnational Corporations*, 2004, 13: 29-55.
- 32 Rasiah, R. and Gachino G. Are Foreign Firms More Productive, and Export and Technology Intensive than Local Firms in Kenyan Manufacturing? 2004, UNU/INTECH Discussion Papers.
- 33 Görg, H. and Strobl E. Multinational Companies and Productivity Spillovers: A Meta Analysis. *Economic Journal*, 2001; 111: 723-739.
- 34 Caves, R. Multinational Firms, Competition, and Productivity in Host-Country Markets. *Economica*, 1974; 41: 176-193.
- 35 Globerman, S. Foreign Direct Investment and Spillovers Efficiency Benefits in Canadian Manufacturing Industries. *Canadian Journal of Economics*, 1979; 12: 42-56.
- 36 Blomström, M. and Persson H. Foreign Investment and Spillover Efficiency in an Underdeveloped Economy: Evidence from the Mexican Manufacturing Industry. *World Development*, 1983, 11: 493-501.
- 37 Nadiri, M. US Direct Investment and the Production Structure of the Manufacturing Sector in France, Germany, Japan, and the UK. 1992, New York University Working Paper.
- 38 Blomström, M. and Wolf E. Multinational Corporations and Productive Convergence in Mexico, 1994 in W. Baumol, R. Nelson and Wolf E. (eds.), *Convergence of Productivity: Cross-National Studies and Historical Evidence*. Oxford: Oxford University Press, 263-283.
- 39 Hurbert, F. and Pain N. (2000) Inward Investment and Technical Progress in the UK Manufacturing Sector. Paris: OECD Economics Department Working Paper.
- 40 Djankov, S. and Hoekman B. Foreign Investment and Productivity Change in the Czech Republic. *World Bank Economic Review*, 2000; 14: 49-64.
- 41 Konings, J. The Effects of Foreign Direct Investment on Domestic Firms: Evidence from Firms Level Panel Data in Emerging Economies. *Economics of Transition*, 2000; 9: 619-633.
- 42 Zukowska-Gagelmann, K. Productivity Spillovers from Foreign Direct Investment in Poland. *Economic Systems*, 2002 (forthcoming).

- 43 Cantwell, J. *Technological Innovation and Multinational Corporations*. 1989, Oxford: Basil Blackwell.
- 44 Kokko, A. *Technology, Market Characteristics and Spillovers*. *Journal of Development Economies*, 1994; 43: 279-293.
- 45 Kokko, A., Tansini R. and Zejan M. *Local Technological Capability and Productivity Spillovers from FDI in the Uruguayan Manufacturing Sector*. *Journal of Development Studies*, 1994; 32: 602-611.
- 46 Girma, S. *Geographic Proximity, Absorptive Capacity and Productivity Spillovers from FDI: A Threshold Regression Analysis*. 2002, GEP Research Paper.
- 47 Girma, S., Greenway D. and Wakelin K. *Who Benefits from Foreign Direct Investment in the UK?* *Scottish Journal of political Economy*, 2001; 48: 119-133.
- 48 Haskel, J., Pereira S and. Slaughter M *Does Inward Foreign Direct Investment Boost the Productivity of Domestic Firms?* 2002, NBER Working Paper.
- 49 Barrios, S., Dimelis, S. Louri H. and Strobl E. *Efficiency Spillovers from the EU Periphery: A Comparative Study of Greece, Ireland and Spain*. 2002, FEDEA Discussion Paper.
- 50 Yudaeva, K., K. Kozlov, N. Melentieva and N. Ponomareva *Does Foreign ownership Matter?* *Economics of Transition*, 2003; 11: 383-409.
- 51 Marin A. and Bell M. *Technology Spillovers from Foreign Direct Investment (FDI): An Exploration of the Active Role of MNC Subsidiaries in the Case of Argentina in the 1990's*. 2003, Paper presented at the DRUID Summer Conference.
- 52 Chudnovsky, D., Lopez A. and Rossi G. *Foreign Direct Investment Spillovers and the Absorption Capabilities of Domestic Firms in the Argentine Manufacturing Sector (1992-2001)*. 2004, MERCOSUR Economic Research Network (forthcoming).
- 53 Girma, S. and Wakelin K. *Regional Underdevelopment: Is FDI the Solution? A Semi Parametric Analysis*. 2001, CEPR Discussion Paper DP2995.
- 54 Ivarsson, I. *Transnational Corporations and the Geographical Transfer of Localized Technology: A Multi-Industry Study of Foreign Affiliates in Sweden*. *Journal of Economic Geography*, 2002; 2: 221-247.
- 55 Blomström, M. *The Economics of International Investment Incentives in OECD*, ed., *International Investment Perspectives*, 2002, Paris: OECD.
- 56 Blomström, M., Kokko, A. and Zejan M. *Foreign Direct Investment: Firm and Host Country Strategies*. 2000, London: Macmillan Press.
- 57 Kugler, M. *The Diffusion of Externalities from Foreign Direct Investment: Theory Ahead of Measurement*. 2000, University of Southampton Working Paper.
- 58 Blalock, G. *Technology From Foreign Direct Investment: Strategic Transfer Through Supply Chains*. 2001, University of California Working Paper.
- 59 Batra, G., Morisset J. and Saggi K. *Vertical Linkages Between Multinationals and Domestic Suppliers: Whom Do They Benefit and Why?* 2003, International Finance Corporation Working Paper.
- 60 Schoors, K. and Van der Tol B. *Foreign Direct Investment Spillovers Within and Between Sector: Evidence from Hungarian Data*. 2002, Ghent University Working Paper.
- 61 Smarzynska Javorcik, B. and Spatareanu M. *Disentangling FDI Spillover Effects: What Do Firm Perceptions Tell US?* (2005) in Moran, H., Graham E. and Blomström M. (eds.) *Does Foreign Direct Investment Promote Development*, Institute for International Economics.
- 62 Lipsey, R. and Sjöholm, F. *Host Country Impacts of Inward FDI: Why Such Different Answers?* 2005 in Moran, H., E. Graham and M. Blomström (eds.) *Does Foreign Direct Investment Promote Development*, Institute for International Economics.
- 63 Lall, S. *Vertical Inter-Firm Linkages in LDCS: An Empirical Study*. *Oxford Bulletin of Economics and Statistics*, 1980; 42: 203-226.
- 64 Rhee, Y.W. and Belot T. *Export Catalysts in Low Income Countries: A Review of Elven Success Stories*. 1990, World Bank Discussion Paper.
- 65 Moran, T. *Parental Supervision: The New Paradigm for Foreign Direct Investment and Development*. 2001, Washington D.C: Institute for International Economics.
- 66 UNCTAD *World Investment Report. Transnational Corporation, Market Structure and Competition*. 1997, New York and Geneva.

- 67 Altenburg, T. Linkages and Spill-overs Between Transnational Corporations and Small and Medium-Sized Enterprises in Developing Countries – Opportunities and Policies. 2000, German Development Institute Working Paper.
- 68 Pavlinek, P. and Smith A. Internationalization and Embeddedness in East-Central European Transition: The Contrasting Geographies of Inward Investment in the Czech and Slovak Republics. *Regional Studies*, 1998; 32: 619-638.
- 69 Pavlinek, P. Regional Development Implications of Foreign Direct Investment in Central Europe. *European Urban and Regional Studies*, 2004; 11: 47-70.
- 70 Mirza, H., Giroud A. and Köster K. Recent Flows in Foreign Direct Investment and Technology Transfer to ASEAN Economies. 2003, University of Bradford Working Paper.
- 71 Mirza, H. and Giroud A. Regional Integration and Benefits from Foreign Direct Investment in ASEAN Economies: The Case of Vietnam. *Asian Development Review*, 2004; 21: 66-98.
- 72 Zemplerová, A. Impact of Foreign Direct Investment on the Restructuring and Growth in Manufacturing. *Prague Economic Papers*, 1998; 7: 329-345.
- 73 Hunya, G. Foreign Investment Enterprises in the Investment process of the Czech Republic, Hungary and Poland. 1997, Vienna Institute for Comparative Economic Studies.
- 74 Eltő, A. The Competitiveness of Hungarian Companies. 2001, Institute for World Economics Working Papers.
- 75 Zemplerová, A. and Jarolim M. Modes of FDI Entry and Firm Performance: The Czech Case. *Transnational Corporations*, 2001; 10: 95-110.
- 76 Rojec, M. Restructuring and Efficiency Upgrading with FDI, 2000 in G. Hunya (eds.), *Integration Through Foreign Direct Investment: Making Central European Industries Competitive*, UK and USA: Edward Elgar, 28-49.
- 77 Rojec, M. The Restructuring of Firms in Foreign Privatizations in Central and Eastern European Countries. *Transnational Corporations*, 2001; 10: 1-24.
- 78 PACEC. Assessment of the Wider Effects of Foreign Direct Investment in Manufacturing in the UK. 1995, Report by PA Cambridge Economic Consultants for Department of Trade and Industry.
- 79 Potter, J., Moore B. and Spires R. The Wider Effects of Inward Foreign Direct Investment in Manufacturing on UK Industry. *Journal of Economic Geography*, 2002; 2: 279-310.
- 80 Crone, M. and Roper S. Local Learning from Multinational Plants: Knowledge Transfers in the Supply Chain. 1999, Northern Ireland Economic Research Centre Working Paper.
- 81 Foreign Investment Advisory Service (FIAS). FIAS Project Impact. 2003, Washington D.C.
- 82 Ferencikova, S. Foreign Direct Investment as a Driver for the Restructuring of a Transitional Economy – Internal and Micro-External Change, 2003 in S. T. Marinova and Marinov M. A. (eds.), *Foreign Direct Investment in Central and Eastern Europe*, UK: Ashgate Publishing Limited, 268-283.
- 83 Hunya, G. Recent Impact of Foreign Direct Investments on Growth and Restructuring in Central European Transition Countries. 2002, Vienna Institute for International Economics Research Report.
- 84 Albanian Institute of Statistics (INSTAT). Publications, 2003: 2004: 2005. Available at [www.instat.gov.al].
- 85 Bank of Albania (BA). Publications, 2004: 2005. Available at [www.bankofalbania.org/].
- 86 Van de Ven, A.H. Suggestions for Studying Strategy Process: A Research Note. *Strategic Management Journal*, 1992; 13: 169-188.
- 87 Sounders, M., Lewis P. and Thornhill A. *Research Methods for Business Students*. 2000, Great Britain: Prentice Hall.
- 88 Yin, R.K. The Case Study Method as a Tool for Doing Evaluation. *Current Sociology*, 1992; 40: 121-137.

SESSION 2:

**INFORMATION TECHNOLOGY
AND
I.T. POLICY**

Teaching and Learning Using VLEs: University Teachers' Approaches to Teaching and their Relationship to Conceptions of Teaching and Learning

Petros Lamer¹, Phil Levy², Iraklis Paraskakis³

¹SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, plameras@seerc.info

²The University of Sheffield, Dept. of Information Studies, Western Bank, Sheffield, S10 2TN, UK, P.Levy@sheffield.ac.uk

³City College, Dept. of Computer Science, 13 Tsimiski Str, 54624 Thessaloniki, Greece, paraskakis@city.academic.gr

Recent research into teaching and learning in higher education has established a close link between teachers' conceptions of teaching and learning, teachers' approaches to teaching and students' learning outcomes. It is possible to characterise teachers' approaches to teaching with motivation and strategy dimensions. Also it is argued that teachers who conceive teaching as transmitting knowledge are more likely to use content – centered approaches to teaching, while those who conceive teaching as conceptual change tend to use learning – centered approaches. Therefore the argument is made that university teachers' preferred approaches to teaching are informed by their conceptions of teaching and learning. An issue that needs further investigation though is whether conceptions of teaching and learning and approaches for teaching are context dependent. It is evident that teachers in higher education are facing a difficult challenge from learning technologies, as they have to renew and develop their model of the teaching and learning process well beyond the traditional transmission model. Subsequently teaching approaches that teachers adopt for using Virtual Learning Environments (VLEs) have to be supported with conceptions of learning as they are described from a constructivist perspective thus to move away from instructivist information transmission mode to a more flexible learning facilitation mode. The paper concludes by suggesting that fundamental changes of the teachers' approaches to teaching using VLEs are unlikely to happen without changes to teachers' conceptions of and approaches to teaching and learning as they are characterised by the research.

Keywords

Approaches to teaching, conceptions of teaching and learning, eLearning, networked learning, teaching and learning using VLEs.

1. Introduction

What university teachers should teach and how should they teach it using Virtual Learning Environments (VLEs)? This question has been investigated from a variety of perspectives, which describe the qualitatively different ways of conceiving and approaching teaching and learning in general. Prosser et al., [1] describe qualitatively different ways of academics' conceptions of learning and teaching; Samuelowicz and Bain [2] describe qualitative differences in approaches to teaching as being *teacher – focused* and *student- focused*. A further paper Trigwell and Prosser [3] argued that qualitative categories of conceptions of teaching and learning were related with approaches to teaching. A certain conception of teaching for example is called 'teaching as assisting students to

change conceptions' while that for a particular approach is labeled a student – focused strategy aimed at students changing their conceptions. Roberts [4] describes a similar conceptualisation from a research undertaken in teachers' conceptions and approaches using the Web which has shown that teachers prefer approaches to teaching, which are influenced by their conceptions of teaching and learning.

Also students' approaches to learning are likely to be influenced by teachers' conception of and approaches for teaching (e.g., [3]). Students' approaches are categorised as deep and surface (e.g., [5]). The motivation associated with a deep approach to learning is to understand ideas and underlying meanings of the particular task assigned whereas in adopting a surface approach the intention is directed to the text itself leading students to view tasks as external impositions that include a reproductive orientation [6]. Given the well-established link between teachers' conceptions of learning, with teachers' conceptions of and approaches for teaching it seems reasonable to argue that there is a relation between teachers' conceptions of learning and teaching with their approaches for teaching. For example if learning is viewed as a conceptual development then the teacher probably holds a *student-centered approach* to teaching, which will normally result in better academic performance. On the contrary if learning is viewed as accumulating more information through processes such as rote learning then the teacher will have a *teacher - centered* approach which will result in poor academic performance.

Moreover current research has showed that the concept of eLearning as defined by the literature is not aligned with university teachers' conceptions of and approaches for teaching and learning using VLEs. Thus, although there is some evidence of student – centredness in teachers' understanding of eLearning, the conception that dominates in teaching and learning with the use of VLEs is for rote learning that includes information transmission. Also the pedagogical approaches that are currently being used are still limited and fragmented [4]. Therefore for eLearning to be widely adopted by university teachers requires that both conceptions of and approaches for teaching and learning and the wider use of VLEs should reflect or aligned with the eLearning definition as articulated by the literature. The reason for such interest in this topic is the belief that teachers' conceptions of and approaches for teaching will, to a large extent define the nature of teaching and learning and the way that university teachers approach their teaching within a VLE.

Britain and Liber [7] regards eLearning as the fundamental way to develop the use of VLEs for teaching and learning context. Other leading researchers in this field (Goodyear, [8], Mayes and Fowler [9] McConell, [10]; Collis and Moonen, [11], Steeples and Jones, [12]; Salmon, [13]; Littlejohn, [14]) argue that effective practice in eLearning is based on the same criteria about effective practice in learning generally. These effective practices are principally student centered for example the achievement of intended learning outcomes, engagement with the learning tasks and motivation. Effectiveness can also be measured in terms of the impact of teachers to the design for learning. This includes the pedagogical approach that is adopted; the planning, supporting, and designing of learning activities structured carefully to promote more effective learning. If this is accepted, then the implications of designing for learning and pedagogy should be highly considered from university teachers for learning and teaching using VLEs.

eLearning suggests the integration of the use of technology with learning practice. In particular Beetham [15] defines eLearning as "learning facilitated and supported through the use of information and communications technology (ICT). Many researchers argue though that this definition is relatively broad and uncontested, as some would limit eLearning more narrowly to networked learning or online learning. The advantage of using eLearning as a broader definition is that "the widest possible range of learning models and modelling procedures can be included for reference." (p:1, [15]). eLearning supports a range of activities which include teaching and learning using a particular kind of electronic media but also face-to-face interaction. Collaboration and co-operation are fundamental elements but also the need for individual reflection and self – direction towards learning constitutes the pedagogic principles of eLearning. Finally the definition refers to sharing and interacting with online resources for collaborative knowledge production (Roberts, [4]). Therefore eLearning is a social process where

students engage in activities and receive feedback from teachers and peers (Goodyear, [8]). For eLearning to be a social process where the concept of connectivity is being promoted between learners, tutors and online resources the design for learning should include certain elements since “the basic idea behind learning design is that people learn better when actively involved in doing something (i.e. are engaged in a learning activity) (p:2, [16]). These elements are:

- The people involved (learner(s), teacher(s) and other supporting roles)
- The sequence of tasks or activities undertaken
- Desired learning outcomes
- Content resources referenced
- Other tools and facilities required.

The sequence of tasks in the general practice of teaching and learning could be probably the lesson plan, “which represents teachers’ understanding of what outcomes are appropriate in a given learning context, and what tasks will help their learners to achieve them (p: 17, [17]). This understanding is consistent with the practitioner’s pedagogical approach for a particular context of learning. Pedagogical approaches are defined by JISC [17] as:

“Distinct but comparable approaches among which practitioners, working in specific context might make an informed choice...approaches to learning involve beliefs, values and cultural norms ...and these operate at the level of content and task design, allocation of resources, forms of dialogue between teachers and learners and definition of desirable outcomes. Neither lesson plans nor individual learning activities are designed without these factors coming into play.” (p:17, [17]).

In this study the pedagogical approaches that could represent those factors of most relevance to the activity of designing for learning and also to be used as a theoretical framework on which learning and teaching using VLEs could be based on draw mainly on Salmon’s [13] five – step model of eLearning, Prosser and Trigwell’s [18] model of the experience of teaching and Kember and Kwan’s [6] teachers’ conceptions of and approaches to, teaching and learning.

2. The Nature of Conceptions (for teaching and learning).

The nature of conception was described primarily as a way of understanding or experiencing the world. The wide variation in the use of the term *conception* amongst researchers has been discussed by Bowden [19]. An early description of the term conception has been introduced by Marton (p: 181, [20]) as a “relation between human beings and the world around them. Bowden [19] argues that the characteristics of the particular relation were not appropriately defined. Recently however the nature of conceptions has been the subject of attention as “the central kind of phenomenon in describing knowledge (p: 163, [21]). General assumptions about the nature of conceptions have been made which are based on the various research studies that have been conducted. These assumptions are:

- Conceptions refer to people’s ways of experiencing a particular aspect of reality.
- Conceptions have certain aspects. These aspects constitute a relation between the subject (the teacher) and the object (the student). These related parts create meaning [21].
- The assumptions about the nature of conceptions made are closely related to assumptions about the nature of knowledge and thinking [21].
- Conceptions are not naturally given entities, neither are they totally subjectively constructed entities[21].

Roberts [4] argues the significance of research into conceptions which allows for classification beyond procedures or behaviors where the changes are numerous and therefore few conceptions may underlie and explain many different behaviors. Pratt [22] conceives conceptions as:

“specific meanings attached to phenomena, which then mediate our response to situations involving those phenomena... we form conceptions of virtually every aspect of our perceived world, and in so doing, use those abstract representations to delimit something from, and relate it to, other aspects of our world.... in effect, we view the world through the lenses of our conceptions, interpreting and acting in accordance with our understanding of the world.” (p: 204, [22])

Research into university teachers’ conceptions of teaching and learning (e.g. Prosser et al., [1]; Samuelowitz and Bain, [2]; Kember, [23, 24]; Entwistle and Walker [24], Murray and MacDonald [25] Martin et al., [26]) showed variation from limited to more complete ways of conceiving what teaching and learning is about. The more complete conceptions involved assisting students change their conceptions of the subject matter, while the limited conceptions were transmission of the subject information or teacher’s understanding. Thus, there is evidence that the way university teachers conceptualise teaching and learning influences their approaches to teaching but also the extent to which their students reach the intended learning outcomes. Marton and Booth [27] contend that this variation from a general to a more concrete meaning could be viewed as a hierarchical relation between the qualitatively different conceptions about a particular phenomenon with the research outcomes themselves. Kember [23] suggests that this hierarchy can be seen as two distinct categories which can be described as:

- Teacher – focused and content oriented (with an emphasis on the reproduction of correct information or teachers’ conceptions of teacher – focused strategy)
- Student – focused and learning – oriented (teachers’ conceptions concerned with conceptual development)

These opposite views are being further discussed by (Trigwell and Prosser [3]; Prosser et al., [1], Kember, [23], Kember and Kwan, [6]).

In analysing the description of the conception of teaching as *encouraging conceptual development and change*, Entwistle and Walker [24] argue that a sophisticated conception such as this underpin effective teaching which include both content and pedagogical knowledge – *pedagogical content knowledge* that has a significant role in teachers’ approaches about teaching and learning. As Entwistle and Walker [24] argues:

"A sophisticated conception of teaching stems from the teacher’s own deep understanding of the subject, but depends on much more. It requires an act of imagination through which the teacher first envisages the subject from the students’ perspective and then devises ways of helping the students across the initial gulf of incomprehension which separates them from the discourse of the discipline or profession.” (p: 343,[24])

Prosser and Trigwell [18] also contest that teachers’ conceptions about teaching and learning are relational in terms of their prior experiences, their perceptions of the teaching context, their approaches to teaching and their teaching outcomes. These relations are framed in terms of a model of the experience of teaching [18] which shows that teachers enter teaching and learning contexts with certain prior conceptions of teaching and learning. The teaching and learning situations that they find themselves in evoke certain aspects of these prior experiences. This interacts with the context to result in certain perceptions of the teaching situation. Certain approaches to teaching and learning are being adopted which are related to teachers’ perceptions of their situation and also are associated with certain outcomes in terms of their experience. For example, university teachers that conceptualise learning as a conceptual change are more likely to adopt a student focused approach that will lead to a more efficient teaching. On the other hand university teachers’ who experience learning as information transmission are more likely to adopt a teacher – focused approach to teaching that will lead to poor academic performance.

Prosser, Trigwell and Taylor [1] explored the conceptions of teaching and learning of science lecturers. Conceptions of teaching ranged from teaching as transmitting concepts of the syllabus to

teaching as assisting students to change their conceptions. Also the same teachers described a range of conceptions of learning as accumulating more information to satisfy external demands and also as a conceptual change to satisfy internal demands.

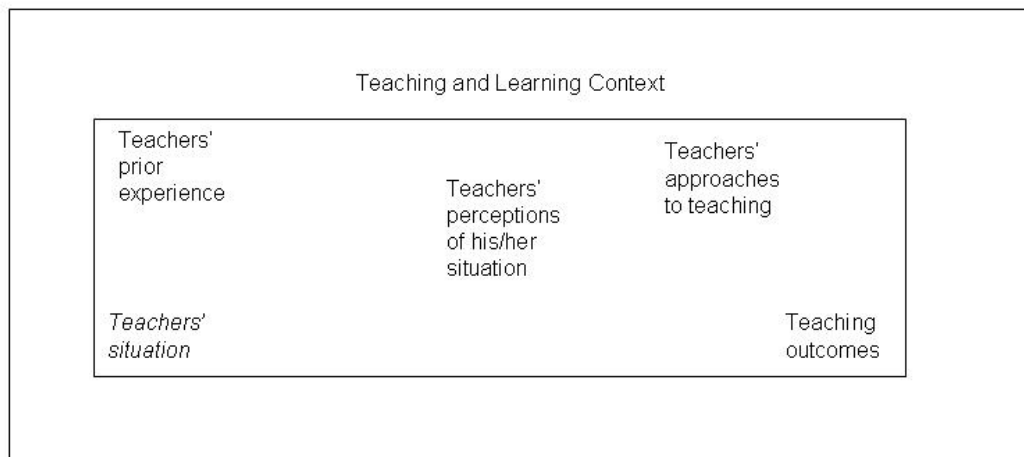


Figure 1 The Model of Teaching Experience (Excerpted from Prosser & Trigwell, [18])

Therefore relating these conceptions on teaching to higher education, it is argued that the qualitative categories of conceptions of teaching and learning are related to those for approaches to teaching and learning (e.g. Trigwell and Prosser, [3]).

3. The Nature of Approaches (to Teaching)

3.1 Motive and Strategy Dimensions

In a phenomenographic study Trigwell and Prosser [3] identified five qualitatively different approaches to teaching as follows:

- Approach A: A teacher – focused strategy with the intention of transmitting information to students;
- Approach B: A teacher – focused strategy with the intention that students acquire the concepts of the discipline;
- Approach C: A teacher/student interaction strategy with the intention that students acquire the concepts of the discipline
- Approach D: A student – focused strategy aimed at students developing their conceptions;
- Approach E: A student – focused strategy aimed at students changing their conceptions.

Approach E, a conceptual change/student – focused approach is one which has the student as the focus of activities. A teacher adopting this approach it matters more what the student learns than what the teacher is doing or covering. The teacher is the one who encourages self – directed learning, who creates the space and the time for students to interact and to discuss the problems of the learning situation, and to develop a *conversation* with students in lectures. Teachers that conceived teaching as an information transmission/teacher focused approach believe that students do not need to be active in the teaching-learning process. As with conceptions of learning these approaches were constituted as hierarchies where the more complete approaches include the more limiting approaches, but not vice versa.

The approaches adopted by the teachers showed to be related to their conceptions of teaching and also to their perceptions of their teaching context. Those teachers who conceive of learning as information

accumulation to meet external demands also conceive of teaching as transmitting information to students, and approach their teaching in terms of teacher-focused strategies. Conversely, those teachers who conceive of learning as developing and changing students' conceptions, conceive of teaching as assisting students to develop and change their conceptions which that indicates that the approach used implies a student – focused way.

Kember [23] argues that the term *teaching approaches* has distinct parallels with *student learning approaches*. These have been characterised as having *motive* and *strategy* components [28] which influenced Prosser and Trigwell [18] to argue that there is a relation between teachers' experiences and their students' experiences in such a way that university teachers who adopt a student focused approach are likely to teach students that adopt a deep approach to their learning while teachers who adopt a teacher – centered approach to teaching are likely to teach students who adopt surface approaches to their study. Therefore the motive dimension could be the teachers' conception of teaching as for instance in Kember and Kwan [6] study on teachers' approaches to teaching the motivational distinction is parallel with the motivational components of student approaches to learning. In other words for the *student - centered* categories the teachers accepted that encouraging and developing student motivation is an important part of the teaching role. Thus, it might logically be assumed that a teacher with a conception of teaching as information transmission it is likely to have a *teacher- centered* approach. Pratt ([29]) identified a similar set of elements “composed of a dynamic and interdependent trilogy of *Actions, Intentions and Beliefs*. A third motive dimension identified in (van Driel, 1997; cited in Entwistle and Walker, [24]) relates to *student-directing* approach. This approach can be viewed as a *blended* approach since teachers' holding that conception believed in teacher control, but also provided opportunities for active learning.

In contrast to these three motives of teachers' approaches to teaching, Roberts [4] found two types of teachers' motives for using the Web for teaching. These are *efficiency and effectiveness*. The efficiency is related to finding ways to do previous teaching and learning activities more cost effectively thus adding greater *efficiency* in teaching and learning. These as Roberts [4] argue stand in contrast to motives that focus on the opportunities the Web brings in terms of eLearning that enhance the *effectiveness* of learning by enabling collaboration, flexibility and self – directed approaches to a particular instance of learning.

Kember and Kwan [6] found five dimensions of the strategy component of the approaches to teaching which can be seen as “continua than discrete categories” (p: 477, [6]). These five dimensions have been examined in terms of a teacher – centered approach and a learning centered approach for conventional teaching. Roberts [4] similarly uses six dimensions of strategy for approaches to teaching campus based students using the Web. If we compare these two studies, we can argue that teachers' approaches in traditional teaching align with teachers' approaches to teaching using the Web. For instance in Kember and Kwan's study, the teacher – centered motive is on marks or on syllabus examination while in Roberts [4] the teacher is absent during students' learning. On the other hand the learning – centered approach for traditional teaching in Kember and Kwan's [6] study is that the teacher encourages the student to discover and construct knowledge while in Roberts [4] student centered approach using the Web, the teacher is actively present and involved in student learning.

3.2 Is the Approach Context Dependent?

Referring back to the model in Figure 1, teachers' approaches to teaching have been widely seen as relational to the teaching and learning context. The approach adopted for a particular instance of learning will depend upon teachers' perceptions of the current situation of the teaching and learning environment. Prosser and Trigwell et al [1] argue that teachers are likely to have preferred approaches to teaching that are influenced by other factors such as curriculum design, presage factors and institutional influences. The relation between teachers' approaches to teaching with the teaching and learning environment might be framed in Kember and Kwan's model (there is an extensive literature on the model itself e.g. Biggs, [5]; and even wider on student approaches to learning and how they are

influenced by the teaching and learning context e.g. Prosser and Trigwell, [18]; Marton et al., [30]) that links conceptions of teaching and learning and approaches to teaching with learning outcomes. The extent to which the context modifies the impact of the current approach that is going to be used by teachers is likely to vary from the following factors:

- An extensive and intensive procedure for course development and approval.
- Intensive procedures for monitoring and reviewing teaching
- Team teaching
- Heavy teaching loads

These factors do not influence teachers' conceptions of teaching rather are likely to influence teachers' approaches to teaching. For that reason Kember and Kwan [6] linked the institutional influence to conceptions of teaching as a dotted line. Therefore as Pratt [29] argues conceptions of teaching are associated more with the beliefs and intentions of the individual than with any particular context. On the other hand approaches to teaching are more likely to be influenced and changed by contextual factors [6].

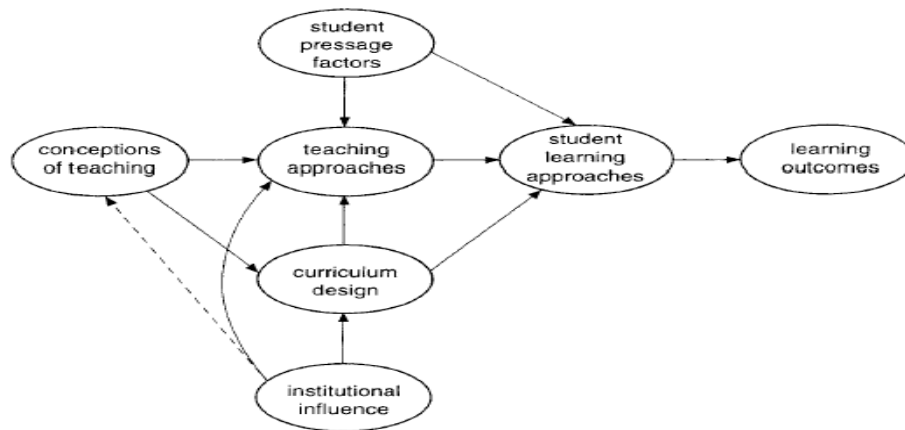


Figure 2 A model linking conceptions of teaching and teaching approaches with learning outcomes (Excerpted from Kember and Kwan, [6]).

4. eLearning

Like any pedagogy, eLearning is based on assumptions about achieving learning outcomes therefore in order to make principled judgments, these assumptions should be clarified. Essentially eLearning is another way of teaching and learning which encompasses all learning undertaken, whether formal or informal, through electronic delivery. However the emphasis as McConnell [10] supports has been “on technology rather than on how technology can facilitate learning, particularly through learner centered and collaborative approaches and the development of skills (p:188, [10]). Thus, McConnell (p:188, [10]) “places an emphasis on collaboration as “a major form of social relationship within a learning context”. The focus is on the learner with the teacher having a facilitative role by assisting students to accomplish the intended learning outcomes. McConnell [10] though prefers the term *networked learning* since the attention is on networking people and resources together. Goodyear [8] supports this argument by defining networked learning as:

“Learning in which communications and information technology is used to promote connections (C&IT): between one learner and other learners and tutors; between a learning community and its learning resources. Such communication can be synchronous and/or asynchronous. It can be text – only multimedia. It may involve learners who are geographically distributed and/or learners who spend much of their time at a common location.” (p:56, [8]).

According to this definition, McConnell [10] argues that the fundamental principles and the

educational potentials of networked learning are:

- The focus on student – student collaboration, where students use the technology to learn from and with each other.
- Students move towards an expectation of sharing, cooperation and collaboration.
- Students construct meaning for themselves from engagement, experience and making sense from their struggle to learn
- Sharing of resources and experience
- Greater degree of openness in the learning process, between the teacher and students, and between students themselves
- In networked learning environments, there is great importance in developing *communities* of learners.

Therefore networked learning is student centered, where students construct meaning through cooperation, collaboration and active participation, the learning is outcome focused and the formation of community of learners is required for encouraging students to develop and share their own situated knowledge and learning resources. The formation of learning communities is further developed by Goodyear's [31] CSALT networked learning model which suggests a distinction between the tasks designed by the tutor and the activities carried out by the learner. Its focus is on collaborative learning thus it demonstrates how learning outcomes can be associated with specific supported learner communities and their activities should be designed in such a way that will result to these learning outcomes. This is supported by Dillenbourg [32] who offers an account as a basis for the analysis of online collaboration where students are engaged in problem – based tasks with a focus on group – based collaborations. Thus the model is useful because of its emphasis on the indirect connection between a collaborative situation and its learning outcomes. The importance of cooperative learning situations for obtaining the intended learning outcomes is emphasised by Goodyear [8] who supports:

“What the learner does is important...we can not influence this directly, but we can try to create situations which are conducive to promoting helpful interactions.”
(p:61, [8])

Roberts [4] argues that an issue for teachers' conceptions would be whether cooperative learning (where individuals work closely with others on their own learning, p12, [10]) could be established through online communication as the definition of eLearning or networked learning implies. This suggests as Britain [16] supports that teachers should focus on the *learning design* with a particular attention on providing effective cooperative learning activities. Britain [16] argues that learning activities are:

Pedagogical techniques that focus on providing activities for learners to perform either in groups or as individuals that help to create deeper, swifter and more effective learning. These may be in the forms of discussions, simulations, mimicry, problem-solving exercises, role plays and quizzes or meta-learning tasks such as construction of mnemonics and mind-maps.” (p3: [16])

Also Britain [16] supports that there are two main advantages associated with the process of designing learning activities. Firstly it provides a framework for teachers to design and structure activities for different learners or group of learners and second these effective designs can be communicated and shared or archived for re-use on future occasions.

The benefits of engaging in the process of learning design are particularly relevant to eLearning where the focus is on content and services at the expend of learning interactions [16]. These benefits could be implemented in a Virtual Learning Environment (VLE) to provide a real - time view where learners are presented with relevant tasks, resources and learning activities for accomplishing the predefined learning outcomes. Most of these systems as Britain and Liber [7] support are intended not to reproduce the traditional classroom environment online, but to use the technology to provide for

learners new tools to facilitate their learning. Therefore the aim of VLEs is to encourage cooperative and resource – based learning and to allow greater sharing and re-use of resources Britain and Liber [7] defines VLEs as:

“Learning management software systems that synthesise the functionality of computer-mediated communications software (e-mail, bulletin boards, newsgroups etc) and online methods of delivering course materials”.(P:1, [7])

The issue for teachers’ approaches to teaching using VLEs would be whether the usage of VLEs for promoting a wider range of learning styles and encouraging more flexible and collaborative learning environments could be aligned and supported by the type of approaches that teachers adopt using a VLE for teaching and learning.

5. Conceptions of and Approaches for Teaching and Learning Using VLEs.

Biggs (p:25, [5]) argues that a good pedagogical design is the one that ensures that there are no inconsistencies between the

- Curriculum that teachers teach
- The teaching methods that teachers use
- The assessment procedures that teachers adopt
- The teaching and learning environment that teachers create for interactions with students
- The institutional climate, the rules and procedures that teachers have to follow.

To achieve that complete consistency there is a need to examine how VLEs are used for promoting what the learner is actually doing. In other words placing the learning and teaching activities (TLAs) to the foreground of teachers’ awareness. Thus, Biggs [5] uses the term *constructive alignment* to indicate that the fundamental assumptions about learning should be based on constructivist theory. The relevant point is that the alignment process cannot proceed without firstly the teachers’ conceptions about learning converge with the constructivist theory, and then adopting teaching approaches that align with those assumptions. The key to explore that is to see how VLEs are used in order to reflect or align either to a content oriented or a student centered conception of, and approach for, teaching and learning.

Salmon [13] developed a five stage model which provides an example of how participants can benefit from working on group based tasks in situations that generate interactions which will lead to the desired learning outcomes. The model also describes the stages of progressing towards successful eLearning both for students and teachers which Salmon [13] refers to them as *e-moderators*. The emphasis is on how to motivate students to build learning from appropriate online tasks through interactive online discussions. The pedagogical approach for using the model provides a framework, which implies a commitment to constructivist tasks that support collaboration with the greatest possible degree of dialogue.

Stage one involves individual access and the induction of participants into eLearning, which are essential prerequisites for online conference participation. Stage two involves individual participants establishing their online identities and then locating others with whom to interact. At stage three participants engage in mutual exchange of information. In that stage a form of cooperation occurs where each student supports the other participant’s goal. At stage four, course related group discussions develop and the discussion becomes more collaborative. Finally, personal development and real reflection will occur on the learning process for achievement of desired learning outcomes at stage five.

The practical implications for applying Salmon’s e-tivities to teaching and learning in higher education are that teachers should have in their focal awareness particular aspects of online teaching. These aspects according to Collins and Berge [33] are:

- Pedagogical, which is based on modeling online behaviour as well as providing insights based on the teacher's subject knowledge.

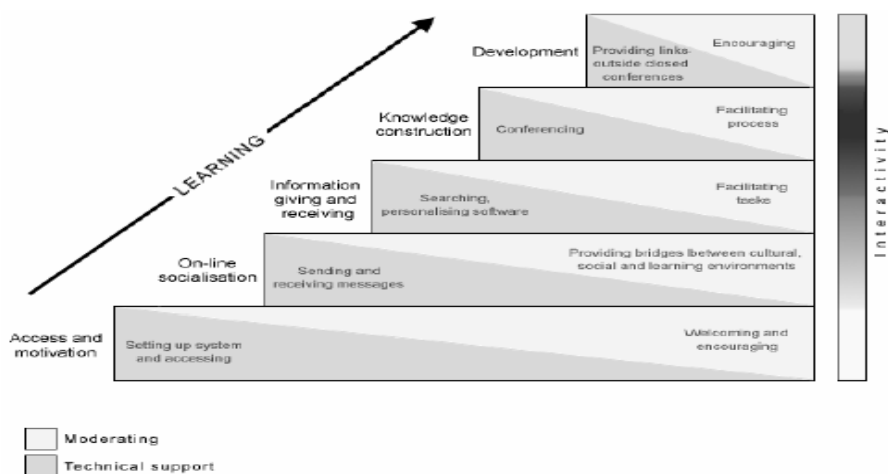


Figure 3 Salmon's e-tivities approach (Excerpted from Salmon, [13])

- The creation of a social environment which includes the recognition to students' contributions and assisting individuals to work together
- Managerial duties that involve setting the pace; visiting and revisiting objectives
- The teacher should be comfortable and proficient with the technology and must ensure that all students are too in order both (teachers and students) to concentrate on the learning tasks.

This implies that teachers' conceptions of and approaches to teaching and learning should be converged and developed from a constructivist perspective in order to enhance learning through dialogue and joint reproduction of knowledge artifacts [34]. This approach will be indicative with the student – centered approach, and will move away from the behavioristic content-based paradigm. In terms of using VLEs this requires that teachers are comfortable and fluent in the use of VLEs in teaching and learning. And most importantly teachers should design learning environments where there is a *negotiation of meaning* and reflection on what has been learned and therefore there will be an understanding of the pedagogical foundations of a new approach (such as the e-tivities model) where the teachers will be the *facilitators of understanding* rather than the *purveyors of knowledge*. Laurillard [35] supports this by arguing that university education should go beyond access to information or content and include engagement with others during the gradual development of their personal understanding. This engagement is developed through interaction between teachers and students and forms the basis of a dialogue-based approach to the education process. Stimulating dialogue among students and between students and tutors has positive effects on social, motivational and attitudinal outcomes thus teachers are required to demonstrate the characteristics and processes identified by Salmon [13], Goodyear [8], Dillenbourg [32], Biggs [5] and McConnell [10].

6. Conclusions (on Teachers' conceptions of and approaches for teaching and learning using VLEs)

Recent research has established a close link between teachers' conceptions of teaching and their conceptions of learning thus it has shown that there is a relationship between teachers' approaches to teaching and their conceptions of teaching and learning that will likely affect teaching and learning outcomes. There is less evidence however of *what conceptions* teachers should have in their focal awareness that will likely determine how approaches are actually formed and developed. Prosser et al., [1] argue that conceptions influence approaches but also as Entwistle and Walker [24] supports these

approaches can influence conceptions in terms of restructuring their awareness of teaching and learning for “encouraging students to reach higher epistemological levels and a deeper understanding of the discipline or professional area” (p:359, [24]). Therefore it would seem reasonably to assume that teachers’ reflections on approaches to using VLEs should converge with principles that promote collaboration, and dialogue where meaningful teaching and learning activities should be in the foreground of teachers’ awareness.

Teachers’ approaches for using VLEs are likely to be linked with teacher focused and content focused categories and are characterised as having motive and strategy components. Therefore we can assume that the teachers who conceive teaching as assisting students to acquire concepts of the syllabus they are likely to have a motive for collaboration and negotiation of meaning, which will result to enhancement of learning and to a student-centered strategy. However Roberts [4] argues that networked learning is more likely to be adopted by teachers whose motivation is learning enhancement rather than imparting information. Consequently teachers’ pedagogic ideas should be developed differently in the new teaching and learning context – that of a VLE –for promoting collaboration, dialogue, negotiation of meaning and most importantly conceptual change.

However two issues have to be explored further for teaching using VLE’s: Firstly how the design for learning should be formed that will provide effective collaborative learning activities for a particular instance of learning and therefore what changes should be made to teachers’ conceptions to reflect this kind of learning design. And secondly to what extent a constructivist perspective could be aligned or supported by the type of approach (teacher focused or student focused) that teachers adopt for teaching and learning using VLEs.

References

- 1 Prosser, M., Trigwell, K., Taylor, P, A Phenomenographic study of Academics' Conceptions of Science Learning and Teaching. *Learning and Instruction*, 1994. 4: p. 217-231.
- 2 Samuelowicz, K., Bain, J, Revisiting academics' beliefs about teaching and learning. *Higher Education*, 2001. 41: p. 299- 325.
- 3 Trigwell, K., Prosser, M, Congruence between intention and strategy in university science teachers' approaches to teaching. *Higher Education*, 1996. 27: p. 1-11.
- 4 Roberts, G., Teaching using the Web: Conceptions and approaches from a phenomenographic perspective. *Instructional Science*, 2003. 31: p. 127-150.
- 5 Biggs, J., *Teaching for Quality Learning at University*. 1999, Buckingham: Society for Research into Higher Education & Open University Press.
- 6 Kember, D., Kwan, Kam-Por, Lecturers' Approaches to Teaching and their Relationship to Conceptions of Good Teaching. *Instructional Science*, 2000. 28: p. 469 - 490.
- 7 Britain, S., Liber, O, A Framework for Pedagogical Evaluation of Virtual Learning Environments, JISC, Editor. 1999. p. Available at www.jisc.ac.uk.
- 8 Goodyear, P., Psychological Foundations for Networked Learning, in *Networked Learning: Perspectives and Issues*, C. Steeples, Jones, C, Editor. 2003, Springer: London.
- 9 Mayes, T., Fowler, C, Learning Technology and Usability. *Interacting with Computers*, 1999. 11: p. 485-497.
- 10 McConnell, D., *Implementing Computer Supported Cooperative Learning*. 2000, London: Kogan Page.
- 11 Collis, B., Moonen, J, *Flexible Learning in a Digital World*. 2001, London: Kogan Page.
- 12 Steeples, C., Jones, C, *Networked Learning: Perspectives and Issues*. 2003, London: Springer.
- 13 Salmon, G., *E-tivities*. 2002, London: Kogan Page.
- 14 Littlejohn, A., *Reusing Online Resources*. 2003, London: Kogan Page.
- 15 Beetham, H., Review: Developing e-Learning Models for the JISC Practitioner Communities. 2004: p. [Available at www.jisc.ac.uk].
- 16 Britain, S., A Review of Learning Design. A Report for the JISC E-learning Pedagogy Programme, 2004: p. [Available at www.jisc.ac.uk].

- 17 JISC, Designing for Learning: An update on the Pedagogy strand of the JISC eLearning Programme. 2004: p. [Available at www.jisc.ac.uk].
- 18 Prosser, M., Trigwell, K, Understanding Learning and Teaching. 1998, Buckingham: Open University Press.
- 19 Bowden, J., The nature of phenomenographic research, in Phenomenographic Research: Variations in Method, J. Bowden, Walsh, E, Editor. 1994, Office of the Director, EQUARD, RMIT: Melbourne. p. 1- 16.
- 20 Marton, F., Phenomenography - Describing Conceptions of the World Around us. Instructional Science, 1981. 10: p. 177-200.
- 21 Svensson, L., Theoretical Foundations of Phenomenography. Higher Education Research and Development, 1997. 16(2): p. 159- 171.
- 22 Pratt, D., Conceptions of teaching. Adult Education Quarterly, 1992. 42(4): p. 203- 220.
- 23 Kember, D., A Reconceptualisation of the Research into University Academics' Conceptions of Teaching. Learning and Instruction, 1997. 7(3): p. 255- 275.
- 24 Entwistle, N., Walker, P, Strategic Alertness and Expanded Awareness within Sophisticated Conceptions of Teaching. Instructional Science, 2000. 28: p. 335-361.
- 25 Murray, K., MacDonald, R, The Disjunction Between Lecturers' Conceptions of Teaching and their Claimed Educational Practice. Higher Education, 1997. 33: p. 331 - 349.
- 26 Martin, E., Prosser, M, Trigwell, K, Ramsden, P, Benjamin, J, What University Teachers Teach and How they Teach it. Instructional Science, 2000. 28: p. 387 - 412.
- 27 Marton, F., Booth, S, Learning and Awareness. 1997, NJ: Lawrence Erlbaum Associates.
- 28 Biggs, J., Student Approaches to Learning and Studying. 1987, Melbourne: Australian Council for Educational Research.
- 29 Pratt, D., Reconceptualising the Evaluation of Teaching in Higher Education. Higher Education, 1997. 34: p. 23 - 44.
- 30 Marton, F., Hounsell, D, Entwistle. N, The Experience of Learning. 1984, Edinburgh: Scottish Academic Press.
- 31 Goodyear, P., Pedagogical Frameworks and Action Research in Open and Distance Learning. CSALT Working Paper, 1999.
- 32 Dillenbourg, P., what do you mean by collaborative learning? in Collaborative learning: Cognitive and computational approaches, P. Dillenbourg, Editor. 1999, Elsevier: Oxford. p. 1-19.
- 33 Collins, M., Berge, Z. Facilitating Interaction in Computer Mediated Online Courses. in FSU/AECT Distance Education Conference. 1996. Tallahassee FL: Available at [<http://www.emoderators.com/moderators/flcc.html>].
- 34 Anderson, T., Annand, D, Wark, N, The Search for Learning Community in Learner Paced Distance Education. Australian Journal of Educational Technology, 2005. 21(2): p. 222-241.
- 35 Laurillard, D., Rethinking University Teaching. 2nd ed. 2002, London: Routledge/Falmer.

Querying and Visualizing the Semantic Web

Georgios Meditskos, Efstratios Kontopoulos, Nick Bassiliades

*Department of Informatics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece,
{gmeditsk, skontopo, nbassili}@csd.auth.gr*

The amount of unstructured information on the Web has reached a critical level, suggesting the need for the Semantic Web (SW). The SW expresses an initiative to improve the World Wide Web, by adding semantic content to the provided information, thus making the data accessible not only to humans but to machines as well. The fulfilment of this goal requires the contribution of a variety of research areas and tools: reasoning techniques that give the opportunity to agents/machines to act autonomously on the environment of the Web in order to fulfil the users' needs and tools for assisting the end users in expressing these needs and comprehending the derived results. This paper presents two systems, namely O-DEVICE and VDR-DEVICE, developed by the Intelligent Systems and Knowledge Processing (ISKP) group¹ at the Department of Informatics, Aristotle University of Thessaloniki, Greece. These two systems are designed to function in the SW environment. O-DEVICE is a system for querying and inferencing about ontologies expressed in the OWL Lite sublanguage and VDR-DEVICE is a visual integrated development environment for developing, using and visualizing defeasible logic rule bases on top of RDF ontologies.

Keywords

Semantic Web, Reasoning System, Visualization Techniques, RuleML, Defeasible Logic.

1. Introduction

The Semantic Web (SW) [0] constitutes an effort to improve the current Web, by targeting at the meaning rather than the presentation of information. This way the content of the Web is made accessible not only to humans, as it is today, but to machines/agents as well, with the latter being now able to understand the available information on the Web and act on behalf of users, in order to achieve their goals in an automated and accurate way.

To describe information appropriately, knowledge representation languages based on XML have been proposed, such as RDF [0] and OWL [0]. Both languages are used to annotate the information in a formal and explicit way by defining ontologies, using classes, properties and instances of classes. The difference between the two languages is the degree of expressiveness they offer. Thus, while RDF is capable of defining only subclass and subproperty relationships between classes and properties respectively, OWL goes a step further by offering a higher degree of expressiveness. Having been built on top of RDF, OWL introduces more relationships between classes, properties and instances, allowing the use of property constraints and Boolean operators (union, intersection, etc) in class definitions.

To fulfil the goal of the SW - a Web understandable by machines - the annotation of information is not enough by itself. There also exists the need for a variety of appropriate tools such as reasoners, graphical editors, visualisation tools and environments that successfully integrate the various technologies and languages involved in such applications. Reasoners are systems able to process SW information and answer queries over that information, so they play a significant role, since they are also able to extract new information that is not stated explicitly, based on the formal semantics of the language used for annotation. The rest of the tools mentioned above are equally important, since they

¹ <http://iskp.csd.auth.gr/>

can greatly assist the end user in exploiting the SW to its full extend as well as for gradually increasing the user trust towards the SW.

In this paper we present two systems: (a) O-DEVICE, a production rule-based system for inferencing and querying ontologies expressed in the OWL Lite language (the simplest sublanguage of OWL), and (b) VDR-DEVICE, a visual integrated development environment for developing, using and visualizing defeasible logic rule bases (rule bases based on defeasible reasoning – an approach to reasoning with incomplete, changing or conflicting information) on top of RDF ontologies. Both systems were developed by the Intelligent Systems and Knowledge Processing (ISKP) group at the Programming Languages and Software Engineering (PLaSE) Laboratory of the Department of Informatics, Aristotle University of Thessaloniki, Greece.

O-DEVICE utilizes an existing production rule system, named CLIPS [0], taking advantage of its efficiency and speed and augments it with an OWL-to-objects mapping mechanism and a deductive query language in order to handle OWL semantics. The system performs inference by following an approach we call Dynamic Rule Generation and it is able to answer queries over the instances of the knowledge base.

VDR-DEVICE, on the other hand, integrates in a user-friendly graphical shell, a defeasible reasoning system that processes RDF data and RDF Schema ontologies, a visual RuleML-compliant rule editor and graph-generating tools for visualizing defeasible logic rule bases. The system supports the multiple rule types of defeasible logic, as well as priorities among rules. It also supports two types of negation (strong, negation-as-failure) and conflicting (mutually exclusive) literals.

Both systems are based on a previous implementation, called R-DEVICE [0], which has been developed by our group too. R-DEVICE is a deductive object-oriented knowledge base system for reasoning over RDF metadata. It is based on an OO RDF data model, different than the established triple-based model, which maps resources to objects and encapsulates properties inside resource objects, as traditional OO attributes. R-DEVICE features a powerful deductive rule language [0] which is able to draw and materialize inference both on the RDF schema and data. The rule language includes features such as ground and generalized path expressions, stratified negation, aggregate, grouping, and sorting functions. The syntax of R-DEVICE rules follows the OPS5/CLIPS paradigm. Furthermore, an XML syntax is provided that extends RuleML [0] and especially the version that supports OO features and negation-as-failure.

The rest of the paper is organized as follows: Section 2 presents the O-DEVICE system more extensively, while section 3 displays the key-features and functionality points of the VDR-DEVICE system. Finally, section 4 concludes this paper and poses directions for future work.

2. O-DEVICE: Inferencing about and Querying OWL Ontologies

The ability to infer new information is a critical characteristic of every reasoning system and defines its completeness and soundness. Unfortunately, there is a tradeoff between scalability, in terms of processing and querying time, and reasoning capabilities. For example, it is almost infeasible to build a complete reasoning system for OWL Full because the great degree of expressiveness does not offer computational guarantees. Thus, most of the reasoning systems aim at the two sublanguages of OWL, OWL Lite and OWL DL that demand fewer reasoning capabilities, especially OWL Lite. Furthermore, the majority of realistic web applications usually involve large number of instances, so a reasoning system should be able to handle efficiently large ABoxes (queries over instances).

The O-DEVICE ([0], [0]) is a system that inferences over (on top of) OWL documents. It exploits the advantages of the object-oriented programming model by transforming OWL ontologies into classes, properties and objects of the OO programming language provided within CLIPS, called COOL.

While most rule-based inference engines utilize rules that operate at the level of triples, in our approach we consume triples to build an object-oriented schema out of an ontology and to materialize OWL instances as objects that encapsulate their properties into their slots. We utilize production rules to infer knowledge based on the ontology and perform appropriate actions in order to materialize the

inferred knowledge on the schema and on the knowledge base. The system also supports a deductive rule language which is used to express queries on the OWL instances. The advantage of this object-oriented approach is that knowledge derived from ontologies is no more scattered in the form of triples across the memory. Instead, rules match a considerable less number of objects (than triples). To make the matching procedure even faster, we introduce a *Dynamic Rule Generation* approach that generates *domain dependent* inference rules that have simpler conditions and thus, faster activation time.

2.1 System Description

The system consists of three basic modules: the *OWL Translator*, the *Query Translator* and the *Result Extractor* (Figure 1). All modules are connected to the CLIPS Production Rule System. The next paragraphs briefly describe each module functionality and role in the overall architecture.

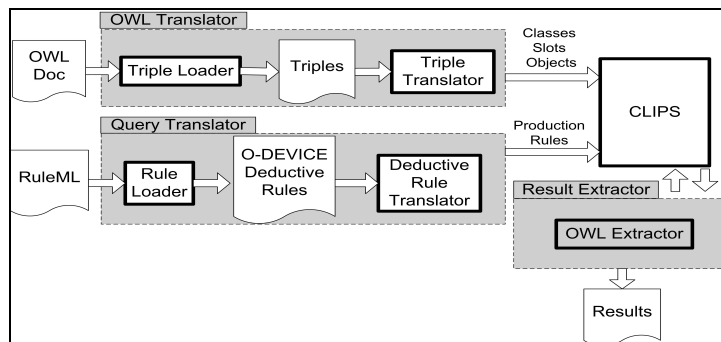


Figure 1 O-DEVICE architecture

OWL Translator. This module is responsible for transforming an OWL document into the object-oriented schema of CLIPS Object-Oriented Language (COOL). It consists of the *Triple Loader* and the *Triple Translator*. The *Triple Loader* loads an OWL document into the system. The document may be stored either locally or remotely, i.e. accessed through a URL. With the aid of the *ARP Parser* [0] the document is transformed into a set of triples which are loaded into main memory and are forwarded to the *Triple Translator*. The latter consists of a set of CLIPS production rules that transforms OWL constructs into COOL object-oriented schema of classes, slots and objects.

Query Translator and Result Extractor. It is responsible for translating incoming queries, which are expressed as deductive rules, into CLIPS production rules. Queries can be expressed either by using a RuleML-like syntax and then transform them into the native CLIPS-like rule language using the *Rule Loader*, or directly into the native syntax. In both ways, the deductive rules are finally passed into *Deductive Rule Translator* which compiles them into CLIPS production rules.

Result Extractor. It is responsible for extracting the results of the executed queries from CLIPS and exporting them into a file using OWL/RDF syntax.

2.2 Building the Object-Oriented Schema

An OWL/RDF document consists of a set of statements in the form of *<subject predicate object>*. Each such statement (triple) denotes the relation that exists between the *subject* and the *object* through the *predicate*. The system reads these triples and materializes the explicit information derived from them according to OWL semantics, using rules and a predefined object-oriented schema in COOL of the OWL built-in classes and properties [0].

Classes. User-defined classes of the ontology are instances of *owl:Class*. For each OWL class such an object (called *meta-object*) is created and the system uses the information stored in them in order to build the actual CLIPS classes (defclass constructs).

Appropriate rules track such objects and build the corresponding class hierarchy (which may need redefinition) based on the values of *rdfs:subClassOf*, *owl:intersectionOf* and *owl:equivalentClass* constructs. If there are no values for these slots, classes become direct subclasses of *owl:Thing*.

Property Instances. These instances define properties in the ontology. Properties could be either instances of *owl:DatatypeProperty* or *owl:ObjectProperty* classes (including their subclasses) and the system generates the objects of the corresponding classes. Rules track such objects and based on the *rdfs:domain* and *rdfs:range* values, create the corresponding slots in the classes. During this procedure, classes may need redefinition if they have already been created.

User-defined class instances. These are the actual objects of user-defined classes in the object-oriented schema. By the time these objects are created, system can populate their slots (user-defined properties) with the values from the ontology instances.

2.3 Extended OWL Semantic Transformations with a Dynamic Rule Generation Technique

In addition to the basic transformations that the system applies at load time in order to build an initial object-oriented infrastructure, it also utilizes production rules to infer information that stems from the semantics of OWL. Since, in our approach, the information is represented using objects and slots, production rules require several joins between the objects of the knowledge base. Thus, the more data we have, the more joins are required, and experiments have shown that increasing number of joins results in an almost exponential increase of load time.

The above observation, led us to adopt an approach based on *Dynamic Rule Generation*. The system generates inference rules based every time on the specific characteristics of the ontology. In that way, rules have only a single condition in their head, avoiding multiple joins and check the minimum possible number of objects. The example that follows demonstrates the advantage of this approach, describing the way that our system handles transitive properties.

In OWL, when a property P is transitive and the pairs (x, y) and (y, z) are instances of P, then it can be inferred that the pair (x, z) is also an instance of P. A general rule for all transitive properties would require three condition elements in its head: one for the transitive property and two more for the objects that participate in the transitive closure. Below we show a simplified rule that implements this.

```
(defrule transitive-property
  (object (is-a owl:TransitiveProperty)(name ?property)
    (rdfs:domain ?domain))
  (object (is-a ?domain)(name ?obj1)(?property $?values1))
  (object (is-a ?domain)(name ?obj2)(?property $?values2))
  (test (belongs ?obj2 $?values1))
=>
  (put-unique-values $?values2 $?values1))
```

Experiments have shown that this kind of rules, i.e rules with more than two condition elements, require much time to be activated, especially when we deal with many objects where the number of joins increases dramatically.

Following the *Dynamic Rule Generation* approach we described, for every transitive property, we generate a rule that grounds as many variables as possible. These rules have limited complexity since they have only one object pattern to match with the minimum number of variables. Consider the following example of a transitive property named *subRegionOf*:

```
<owl:Class rdf:ID="Region"/>
<owl:TransitiveProperty rdf:ID="subRegionOf">
  <rdfs:domain rdf:resource="#Region"/>
  <rdfs:range rdf:resource="#Region"/>
</owl:TransitiveProperty>
<Region rdf:ID="region1" >
  <subRegionOf rdf:resource="#region2"/>
</Region>
```


Region1 is a *subRegionOf* *region2* and *region2* is a *subRegionOf* *region3*. Because *subRegionOf* is a transitive property, the system must infer that *region1* is also a *subRegionOf* *region3*. The resulting rule for the above property following the *Dynamic Rule Generation* approach can be seen below.

```
(defrule gen1
  (object (is-a Region)(name ?obj1)
   (subRegionOf $? ?obj2 $?))
  (test (transitive-closure ?obj1 ?obj2 subRegionOf)
 =>
   (put-unique-values ?obj2 ?obj1))
```

This rule checks every instance value of the *subRegionOf* property of the instance *?name* and in the body of the rule computes the transitive closure. In that way, only instances of the class *Region* and its subclasses are checked, without performing any join in order to calculate the transitive closure, speeding-up rule activation time.

2.4 The Deductive Rule Language of O-DEVICE

The deductive rule language of O-DEVICE supports inferencing over OWL instances represented as objects and defines materialized views over them, possibly incrementally maintained. The conclusions of deductive rules represent derived classes, i.e. classes whose objects are generated by evaluating these rules over the current set of objects. Furthermore, the language supports recursion, stratified negation, path expressions over the objects, generalized path expressions (i.e. path expressions with an unknown number of intermediate steps), derived and aggregate attributes [0]. Each deductive rule in O-DEVICE is implemented as a CLIPS production rule that inserts a derived object when the condition of the deductive rule is satisfied.

The following example shows a rule that retrieves the names of all *Woman* instances that have a value less than 22 in the age property by deriving instances of class *young-woman* with the value *?fname* in the *fname* property:

```
(deductiverule young-women
  (test:Woman (test:age ?x&:(< ?x 22)) (test:fname ?fname))
=>
  (young-woman (fname ?fname)))
```

3. VDR-DEVICE: Developing and Visualizing Defeasible Logic Rule Bases

Defeasible reasoning [0], a member of the non-monotonic reasoning family, represents a simple rule-based approach to reasoning with incomplete, changing and conflicting information. The main advantages of this approach are enhanced representational capabilities coupled with low computational complexity.

Defeasible reasoning can represent facts, rules and priorities and conflicts among rules. Nevertheless, although defeasible logic is certainly a very promising reasoning technology, the development of rule-based applications for the SW can be greatly compromised by two factors. First, the RuleML syntax of defeasible logic, which is briefly exhibited in a next section, is certainly too complicated for an end-user language, creating the need for user-friendly authoring tools. Furthermore, the solid mathematical formulation that forms the basis of defeasible reasoning may seem too complicated. In this case a graphical trace and an explanation mechanism would also be very beneficial.

VDR-DEVICE is a visual integrated development environment for developing, using and visualizing defeasible logic rule bases on top of RDF ontologies. The system integrates in a user-friendly graphical shell, a defeasible reasoning system that processes RDF data and RDF Schema ontologies, a visual RuleML-compliant rule editor and graph-generating tools for visualizing defeasible logic rule bases.

3.1 Defeasible Reasoning – Basic Principles

The main building block of defeasible reasoning is the *defeasible theory* D (i.e. a knowledge base or a program in defeasible logic), which consists of three basic ingredients: a set of facts (F), a set of rules (R) and a superiority relationship ($>$). Therefore, D can be represented by the triple $(F, R, >)$.

In defeasible logic, there are three distinct types of rules: strict rules, defeasible rules and defeaters.

- *Strict rules* are denoted by $A \rightarrow p$ and are interpreted in the typical sense: whenever the premises are indisputable, then so is the conclusion. An example of a strict rule is: “Penguins are birds”, which would become: $r_1: \text{penguin}(X) \rightarrow \text{bird}(X)$.
- Defeasible rules, contrary to strict rules, can be defeated by contrary evidence and are denoted by $A \Rightarrow p$. Examples of defeasible rules are $r_2: \text{bird}(X) \Rightarrow \text{flies}(X)$, which reads as: “Birds typically fly” and $r_3: \text{penguin}(X) \Rightarrow \neg \text{flies}(X)$, namely: “Penguins typically do not fly”.
- *Defeaters*, denoted by $A \sim p$, cannot actively support conclusions, but can only prevent some of them. In other words, they are used to defeat some defeasible conclusions by producing evidence to the contrary. An example of such a defeater is: $r_4: \text{heavy}(X) \sim \neg \text{flies}(X)$, which reads as: “Heavy things cannot fly”. This defeater can defeat the (defeasible) rule r_2 mentioned above.

Finally, the *superiority relationship* among the rule set R is an acyclic relation $>$ on R , that is, the transitive closure of $>$ is irreflexive. Superiority relationships are used, in order to resolve conflicts among rules. For example, given the defeasible rules r_2 and r_3 , no conclusive decision can be made about whether a penguin can fly or not, because rules r_2 and r_3 contradict each other. But if the superiority relationship $r_3 > r_2$ is introduced, then r_3 overrides r_2 and we can indeed conclude that the penguin cannot fly. In this case rule r_3 is called *superior* to r_2 and r_2 *inferior* to r_3 .

3.2 Rule Editor – Design and Functionality

VDR-DEVICE supports a RuleML-compatible [0] syntax, which is the main standardization effort for rules on the SW. Figure 2 displays a fragment of the VDR-DEVICE RuleML-compatible syntax that represents rule r_1 mentioned in the previous section.

```
<imp>
  <_rlab ruleID="r3" ruletype="defeasiblerule" superior="r2">
    <ind>r3</ind>
  </_rlab>
  <_head>
    <neg>
      <atom>
        <_opr> <rel>flies</rel> </_opr>
        <_slot name="name"> <var>X</var> </_slot>
      </atom>
    </neg>
  </_head>
  <_body>
    <atom>
      <_opr> <rel>penguin</rel> </_opr>
      <_slot name="name"> <var>X</var> </_slot>
    </atom>
  </_body>
</imp>
```

Figure 2 A strict rule, written in the RuleML-compatible language of VDR-DEVICE.

Figure 2 clearly demonstrates that the RuleML syntax can be too complex for the end user. Thus, the need for authoring tools that assist end-users in writing and expressing rules is imperative. VDR-DEVICE is equipped with DRREd (Figure 3), a visual rule editor that aims at enhancing user-friendliness and efficiency during the development of VDR-DEVICE RuleML documents [0]. Its implementation is oriented towards simplicity of use and familiarity of interface. And its key features

include: (a) functional flexibility - program utilities can be triggered via a variety of overhead menu actions, keyboard shortcuts or popup menus, (b) improved development speed - rule bases can be developed in just a few steps and (c) powerful safety mechanisms – the correct syntax is ensured and the user is protected from syntactic or RDF Schema related semantic errors.

The rule base is displayed in XML-tree format, one of the most intuitive means of displaying RuleML-like syntax because of its hierarchical nature. The user can traverse the tree and can add to or remove elements from the tree. However, since each rule base is backed by a DTD document, potential addition or removal of tree elements has to obey to the DTD limitations. Therefore, the rule editor allows a limited number of operations performed on each element, according to the element's meaning within the rule tree.

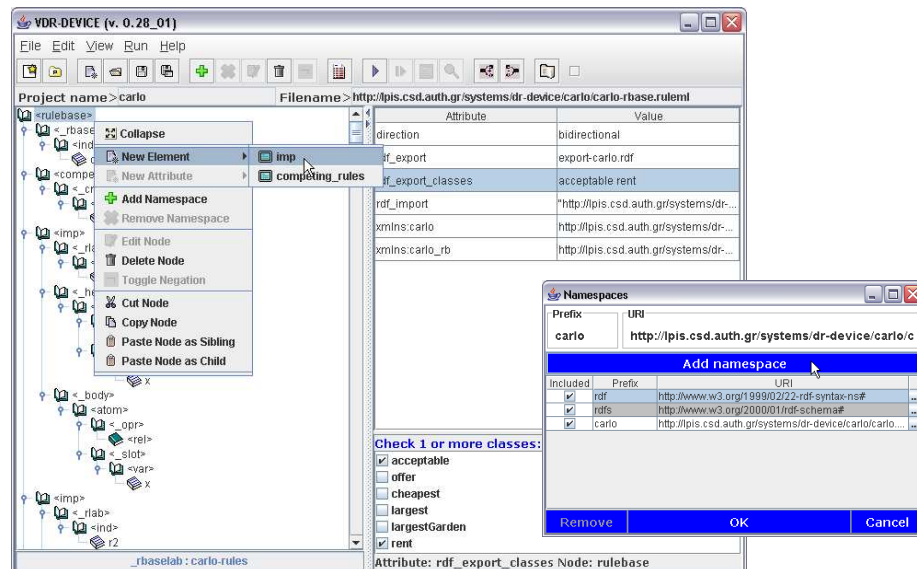


Figure 3 The graphical rule editor and the namespace dialog window.

By selecting an element from the tree, the corresponding attributes are displayed each time. The user can also perform editing functions on the attributes, by altering the value for each one of them. However, the values that the user can insert are obviously limited by the chosen attribute each time.

3.3 The Reasoning System - Architecture and Functionality

The core reasoning system of VDR-DEVICE is DR-DEVICE [0] and consists of two primary components (Figure 4): The *RDF loader/translator* and the *rule loader/translator*. The user can either develop a rule base with the help of the rule editor described previously, or he/she can load an already existing one, probably developed manually. The rule base contains: (a) a set of rules, (b) the URL(s) of the RDF input document(s), which is forwarded to the RDF loader, (c) the names of the derived classes to be exported as results and (d) the name of the RDF output document.

The rule base is then submitted to the *rule loader* which transforms it into the native CLIPS-like syntax through an XSLT stylesheet and the resulting program is then forwarded to the *rule translator*, where the defeasible logic rules are compiled into a set of CLIPS production rules. This is a two-step process: First, the defeasible logic rules are translated into sets of deductive, derived attribute and aggregate attribute rules of the basic deductive rule language, using the translation scheme described in [0]. Then, all these deductive rules are translated into CLIPS production rules according to the rule translation scheme in [0].

Meanwhile, the *RDF loader* downloads the input RDF documents, including their schemas, and translates RDF descriptions into CLIPS objects, according to the RDF-to-object translation scheme in [0], which was also described in section 2.

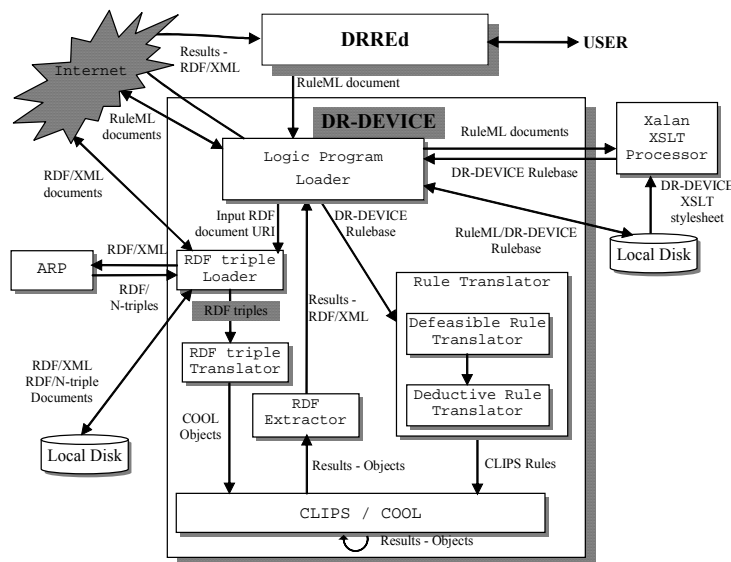


Figure 4 The architecture of the reasoning system.

The inference engine of CLIPS performs the reasoning by running the production rules and generates the objects that constitute the result of the initial rule program. The compilation phase guarantees correctness of the reasoning process according to the operational semantics of defeasible logic. Finally, the result-objects are exported to the user as an RDF/XML document through the RDF extractor. The RDF document includes the instances of the exported derived classes, which have been proved.

3.4 Visualizing a Defeasible Logic Rule Base

DRRed is equipped with a utility that allows the creation of a directed rule graph from the defeasible rule base developed by the editor [0]. This way, users are offered an extra means of visualizing the rule base, besides XML-tree format described above and, thus, possess a better aspect of the rule base displayed and the inference results produced.

More specifically, in order for the desired graph to be constructed, certain rule base elements have to be collected, following the path described in the following paragraphs. For every class in the rule base (i.e. classes that lie at rule bodies and heads) a *class box* is constructed, which is simply a container. The class boxes are populated with one or more *class patterns* during the development of the rule base. For each atomic formula inside a rule head or body, a new class pattern is created and is associated with the corresponding class box. In practice, class patterns express conditions on instances of the specific class.

Each class pattern is visually represented by a rectangle, separated in two adjacent “*atomic formula boxes*”, with the upper one of them representing the positive atomic formula and the lower one representing the negative atomic formula. This way, the atomic formulas are depicted together clearly and separately, maintaining their independence.

Class patterns are populated with one or more *slot patterns*. Each slot pattern consists of a slot name and, optionally, a variable and a list of value constraints. The variable is used in order for the slot value to be unified, with the latter having to satisfy the list of constraints. In other words, slot patterns represent conditions on slots (or class properties).

An example of all the above can be seen in Figure 5. The figure illustrates a class box that contains three class patterns applied on the *person* class and a code fragment matching the third class pattern, written in the RuleML-like syntax of VDR-DEVICE. The first two class patterns contain one slot pattern each, while the third one contains two slot patterns. As can be observed, the argument list of each slot pattern is divided into two parts, separated by ”|”; on the left all the variables are placed and

on the right all the corresponding expressions and conditions, regarding the variables on the left. In the case of constant values, only the right-hand side is utilized; thus, the second class pattern of the box in Figure 5, for example, refers to all the *male* persons. This way the content of the slot arguments is clearly depicted and easily comprehended.

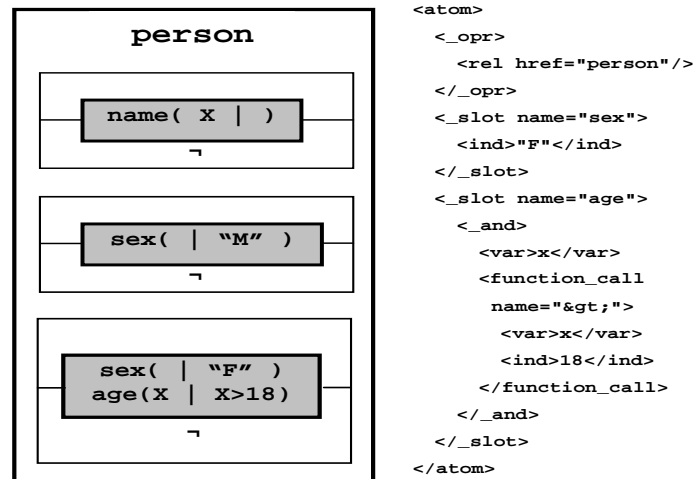


Figure 5 A class box example and a code fragment for the third class pattern

A plan has to be drawn, concerning the location of the elements in the graph-drawing panel of VDR-DEVICE. For this affair, a variation of the rule stratification algorithm found in [0] was implemented. According to the algorithm, the graph elements are placed in *strata*; each stratum is considered as a *column* in the graph drawing panel of the system. Thus, the first stratum is mapped to the first column on the left, the second stratum to the column on the right of the first one and so on. The algorithm aims at giving a left-to-right orientation to the flow of information in the graph; i.e. the arcs in the digraph are directed from left to right, making the derived graph less complex, more comprehensible and easily readable.

The two aspects of the rule base, namely the XML-tree and the directed graph are interrelated, meaning that traversal and alterations in one will also be reflected in the other. So, if for example the user focuses on a specific element in the tree and then switches to the digraph view, the corresponding element in the digraph will also be selected and the data relevant to it displayed.

Figure 6 displays the output of the VDR-DEVICE rule graph drawing module, when a small defeasible logic rule base is loaded. The specific rule base includes four rules (rules r_1 - r_4), all of them are defeasible and the first one “lands” on a positive conclusion (the upper atomic formula box of the head class pattern), while the other three “land” on a negative one (the lower atomic formula box of the head). Furthermore, there exist one base and one derived class. The total number of strata in this case is three (i.e. three columns).

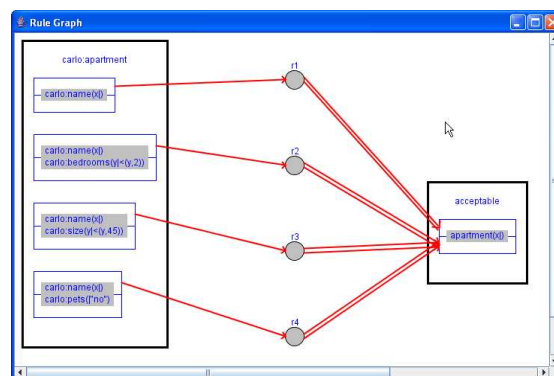


Figure 6 The rule graph drawing module of DRRED

4. Conclusions and Future Work

The Semantic Web is a very promising technology that aims at enhancing the current Web with metadata, thus making its content accessible to machines besides humans. In this paper we presented two systems, O-DEVICE and VDR-DEVICE, developed specifically for the environment of the Semantic Web. The former inferences over OWL metadata, exploiting the advantages of the object-oriented programming model by transforming OWL ontologies into CLIPS classes, properties and objects. The latter constitutes an integrated environment for the development and visualization as well as the reasoning over defeasible logic rule bases on top of RDF ontologies. Both systems, each in its own way, attempt to narrow the gap between the SW and the end user, by providing means that assist in exploiting the potential of the SW to its full extend.

For the future, we intend to augment the O-DEVICE semantic translation procedure in order to handle more OWL constructs, such as *owl:sameAs*. Furthermore, we plan to extend the system to be able to handle ontologies expressed in the OWL DL language, a more expressive sublanguage of OWL. Our final goal is to develop a Semantic Web Service composition system, where services will be described using OWL-S while the composed services will be described using logic rules.

As for VDR-DEVICE, it is planned to be extended soon with a graphical RDF ontology and data editor that will comply with the user-interface of the RuleML editor. Furthermore, we plan to delve deeper into the proof layer of the Semantic Web architecture by enhancing further the rule representation utility demonstrated with rule execution tracing, explanation, proof exchange in an XML or RDF format, proof visualization and validation, etc. These facilities would be useful e.g. for automating proof exchange and trust among SW agents and for eventually increasing the trust of users towards the SW.

The ultimate goal is naturally to merge the two systems into one, by having a single visual environment for developing and deploying rule bases of varying expressiveness and functionality in the SW for a variety of ontology and metadata languages. Further development for the integrated tool would be the persistent storage of metadata in a SW repository, such as Sesame [0] or ICS-FORTH RDF Suite [17] and its deployment as a Semantic Web service.

References

- 1 Berners-Lee T, Hendler J, Lassila O. The Semantic Web. Scientific American, 2001; 284(5), 34-43.
- 2 Resource Description Framework, <http://www.w3.org/RDF/>
- 3 Web Ontology Language (OWL), <http://www.w3.org/2004/OWL/>
- 4 CLIPS 6.23 Basic Programming Guide, <http://www.ghg.net/clips>
- 5 Bassiliades N, Vlahavas I. R-DEVICE: An Object-Oriented Knowledge Base System for RDF Metadata. Int. Journal on Semantic Web and Information Systems, 2(2), pp. 24-90, 2006.
- 6 Boley H, Tabet S, Wagner G. Design Rationale of RuleML: A Markup Language for Semantic Web Rules, Proc. Int. Semantic Web Working Symp., pp. 381-402, 2001.
- 7 Meditskos G, Bassiliades N. Towards an Object-Oriented Reasoning System for OWL, Int. Workshop on OWL Experiences and Directions, B. Cuenca Grau, I. Horrocks, B. Parsia, P. Patel-Schneider (Ed.), 11-12 Nov. 2005, Galway, Ireland, 2005.
- 8 Meditskos G, Bassiliades N. O-DEVICE: An Object-Oriented Knowledge Base System for OWL Ontologies, Proc. 4th Hellenic Conference on Artificial Intelligence (SETN-06), G. Antoniou, G. Potamias, D. Plexousakis, C. Spyropoulos (Ed.), Springer-Verlag, LNAI 3955, pp. 256-266, Heraklion, Crete, 18-20 May, 2006.
- 9 McBride B. Jena: Implementing the RDF Model and Syntax Specification, Proc. 2nd Int. Workshop on the Semantic Web, 2001.
- 10 OWL Web Ontology Language Reference, <http://www.w3.org/TR/owl-ref/>
- 11 Nute D. Defeasible Reasoning. Proc. 20th Int. Conference on Systems Science (pp. 470-477). IEEE Press. 1987.

- 12 Bassiliades N, Kontopoulos E, Antoniou G. A Visual Environment for Developing Defeasible Rule Bases for the Semantic Web. Proc. International Conference on Rules and Rule Markup Languages for the Semantic Web (RuleML-2005), A. Adi, S. Stoutenburg, S. Tabet (Ed.), Springer-Verlag, LNCS 3791, pp. 172-186, Galway, Ireland, 2005.
- 13 Bassiliades N, Antoniou G, Vlahavas I. A Defeasible Logic Reasoner for the Semantic Web. Int. Journal on Semantic Web and Information Systems, 2(1), pp. 1-41, 2006.
- 14 Kontopoulos E, Bassiliades N, Antoniou G, Visualizing Defeasible Logic Rules for the Semantic Web. 1st Asian Semantic Web Conference (ASWC'06) (to be presented), Beijing, China, 3-7 September, 2006.
- 15 Ullman J D. Principles of Database and Knowledge-Base Systems, Vol 1, Computer Science Press, 1988.
- 16 Sesame metadata repository, <http://www.openrdf.org/>
- 17 The ICS-FORTH RDFSuite: High-level Scalable Tools for the Semantic Web, <http://www.ics.forth.gr/isl/r-d-activities/rdf-gr.htm>

A Model Framework to characterise the NEHRS Preparedness: Validation Phase

Leonidas Orfanidis¹, Barry Eaglestone², Panagiotis Bamidis³

¹*South East European Research Centre, 17 Mitropoleos Str, 54624, Thessaloniki, Greece, lorfanidis@seerc.info*

²*University of Sheffield, Dept. of Information Studies, Western Bank, Sheffield, S10 2TN, UK, B.Eaglestone@sheffield.ac.uk*

³*Aristotle University of Thessaloniki, Medical Informatics Laboratory, School of Medicine, Greece, bamidis@med.auth.gr*

This paper addresses the issues of the National Electronic Health Record Systems (NEHRS) development. A model framework is stepwise produced, that allows for the characterisation of the preparedness and the readiness of a country to develop an NEHRS. The model is built from secondary data of published reports, considered within a sample of five developed countries. The model is instantiated for the case of Greece, Romania and the UK, as an example of its first application. Finally, the model is validated using the Delphi method which identifies a number of changes on the model.

Keywords

National Electronic Health Record Systems, Delphi method.

1. Introduction

The development of National Electronic Health Record Systems is still in progress in a number of countries across the world. None of the countries has finished the development of such a system since they face many obstacles which cause serious delays on the national programmes. A large number of reports analyse the factors which seem to generate these difficulties on the NEHRS programmes. This paper studies these factors creating a model which is referred to as Characterisation Model or CM, and it is derived to characterise the readiness of nations to avoid the widely known problems and develop a NEHRS. In addition, in this paper we describe the validation of the model. The synoptic purpose of the paper is the presentation of the CM and the analysis of its validation process referring also on the validation results and conclusions.

The CM has been created as an instrument for assessing national readiness for the development of a NEHRS on the basis of the experience of nations with advanced NEHRS strategies, as a part of a general methodology for applying evidence based know-how in the development of NEHRS strategies (see figure 1). The CM has been derived inductively from secondary data, mainly governmental and related reports on NEHRS strategy (see section 2). In total 134 reports have been studied and analysed using the hermeneutics method [1]. This method presupposed the study of the reports for attaining a general sense of the information and then it followed a more detailed study and analysis classifying all the material into categories. The construction of the CM was from quantified parameters based on those factors which emerged from this study, as appearing to be major influences on such strategies.

Moreover, this paper presents the rationale and design of the method by which the CM has been validated, together with the validation results. Validation of a model requires establishing the extent to which its properties and behavior correspond to those of the phenomena that it attempts to model. However, validation of the CM is problematic, since currently there is no hard data or established theory against which to test this correspondence. Specifically, as yet there is no available hard data

relating to developed NEHRS since they are currently an aspiration rather than a reality. Therefore an alternative methodology is necessary to validate the CM. In this paper we document the method we have selected and present a rationale that establishes its appropriateness and feasibility for our case (model). It is called Delphi method and it is widely known on researchers which are trying to predict the “success” (in terms of providing good quality of results) of new models, theories and methodologies [2]. The Delphi method is relevant when there is no available determinate answers (eg., hard data or well established theory), as is the case in this study. In the absence of other more concrete criteria against which to establish the validity of a model, the Delphi tests the predictions generated by a model against the consensus opinion of experts who have relevant information about the topic of concern. Accordingly, in this form of validation, the model is tested and elaborated against the consensus opinions of a panel of experts within the field of NEHRS strategy. Delphi method is based on a structured process for collecting and distilling knowledge from a group of experts using a series of questionnaires [3]. The key for the success of this validation method was the selection of the group of experts on the area of NEHRS (see Section 3).

The remaining sections of this paper are structured as follows. Section 2, called “Characterisation Model” presents the model aspects and their unit metrics. Section 3, defines and discusses the adaptation of the Delphi methodology used in this validation. The implementation of the methodology is then presented. The problem definition of this research and the selection of the experts’ panel are then defined and explained in Section 4 and 5 respectively. Section 6, describes the design of the questionnaire, which was used as the research instrument for data collection. The final part of the paper presents the results of this validation exercise. An analysis of the data collected in this way is presented in Section 7 followed by final concluding remarks.

2. Characterisation Model

A model framework definition is attempted in order to characterise national preparedness for development of an NEHRS, as derived primarily from governmental reports concerning NEHRS programmes of the sample of developed countries. Thus, the main inputs were the concerns and barriers, apparent in this literature, and faced during the development and application of NEHRS strategies, by the UK, Australia, Canada, the USA and New Zealand.

The model takes the form of a tuple, the elements of which are what emerged as the key factors, which appear to enable and/or impede NEHRS strategy. A particular feature of this model is our attempt to quantify and aggregate those factors so as to deliver a high level quantitative basis for evaluating and comparing nations. Thus, rather than seeking to elaborating the model to capture what is clearly a rich and complex problem area, we have sought to simplifying the situation, so as to deliver a broad measure with which a nation’s NEHRS preparedness can be assessed and compared.

2.1 Important aspects defined by 5 developed countries

Standards (A1)

The development, adoption and enforcement of standards to cover all aspects of NEHRS emerged strongly as a necessary prerequisite to an effective NEHRS. In fact, this factor was a main theme in the NEHRS strategy documents of all of the nations in the sample. Key categories and the associated standards, which emerged from the publications reviewed, are tabulated in Table 1.

This national concern for relevant healthcare information and related standards is well founded. Enforcement is clearly critical to an effective NEHRS, for example, to ensure effective data transfer, storage, security, and accuracy. In particular, an NEHRS is inherently heterogeneous and distributed, and therefore standards are necessary to achieve interoperability between component systems. Certainly, the lack of such standards will lead to a failure during the implementation phase for reasons of incompatibility between system components. At a more general human level, standards are necessary, to achieve a common and correct interpretation of healthcare data. For example, this can be achieved only by enforcing consistent and meaningful use of medical terminologies, imposing a

system of unique identifiers for patients; enforcing protocols and procedures to ensure data quality; while guarding against inappropriate data access through protocols to ensure EHR privacy and security.

In the model of this paper, it is attempted to quantify this factor as a measure of the scope of the standards adopted by a nation. To achieve this, Table 1 provides a list of key standards identified in the documentation reviewed, and categorised into critical areas also cited in the documents. The metric is then computed as the number of categories within which a nation has or plans to adopt one or more of those key standards. The minimum grade of this metric is 1 where no standards from none of the categories or from only one category has or is planned to be adopted. The maximum grade of this unit metric is 7 where at least one of the standards from each category has or is planned to be adopted. The intention is that the result of this calculation will give some indication of the infrastructure (readiness) of each country with respect to health informatics standards and conversely the extent of those areas not covered by nationally adopted standards.

Category of Standards	Name of the Standard
Medical Terminologies and Classification Standards	SNOMED Read Clinical Terms UMLS GALEN MEDCIN CPT-4 LOINC ICPC-2 ICD-10
Message format standards	HL7 CDA CEN/TC 251 ENV 13606-Part 4
Digital image standards	DICOM
EHR content and structure	OpenEHR ASTM HL7 RIM CEN/TC 251 ENV 13606-Part 1,2,3
Secure exchange, storage and access standards	EDIFACT EbXML e-GIF
Unique identifiers standards	Unique identifiers
Member of ISO/TC215 or CEN/TC215	ISO/TC215 and CEN/TC215

Table 1 Categories of the necessary health standards for the development of an NEHRS [4]

Health network architecture (A2)

The type of established or planned network facility determines to a large extent the means by which an NEHRS must capture, store and communicate patient-related health information. Specifically, the network infrastructure facilitates or impedes interoperability between relevant information systems, and determines levels of data security that can be achieved, as well as, the ease of access. The national ICT infrastructure seems to play a key role in the selection of the appropriate technologies for the health network architecture. In particular, the availability of broadband Internet access across the country, together with the computer literacy of citizens and healthcare personnel appear to be an important factor in the selection of appropriate NEHRS solution, since these solutions must be adapted to the needs of the users and the capabilities of the national infrastructure. This relationship between the national health network architecture and the national ICT infrastructure emerged as a major theme, for example, in the following key NEHRS strategy documents.

In particular, interconnection of all healthcare systems within a country, in both public and private health sectors, appears to be an aspiration for developed countries with a legacy of existing heterogeneous health information systems. In order to model and quantify this aspect within the model framework, five measures have been identified, each of which are available from statistics periodically collected by the World Bank [5]. The selected metrics are [4]:

- **A2.1:** Access to Broadband Internet (through DSL or cable modem), measured on a scale of 1 to 7, with 1 denoting no broadband access and 7 denoting wide availability. The World Bank uses a complicated function in order to provide this metric from 1 to 7.
- **A2.2:** The number of telephone lines per 1000 citizens. In order to achieve normalisation of the metric the 1000 is uniformly divided into 7 equal ranges, in order to keep the same normalisation with other metrics of the model (i.e. metric A2.2 ranges from one (1) (1 - 142,9) to seven (7) (857.4-1000). The same grade normalisation is applied in the next three sub metrics.
- **A2.3:** The number of personal computers owned per 1000 citizens. (scale from 1 to 7)
- **A2.4:** The number of Internet users per 1000 citizens. (1 to 7 scale)
- **A2.5:** The number of mobile phones per 1000 citizens. (1 to 7 scale)

Legislation Framework (A3)

A legislation framework is necessary for the provision of data privacy. Regulations and laws are necessary in order to define the kind of access that each kind of user will have. Moreover, legislations are necessary for the supervision of the personal health data.

The laws for health data privacy are the same for the electronic commerce, digital signatures and consumer protection. The World Bank is a notable data statistics source and well established worldwide for the quality of its data. For legislation, they annually provide measures for the efficacy of the laws on the above areas of electronic commerce, consumer protection and digital signatures, under the specific term “Laws related to ICT use”. These laws are measured on a scale of 1 (minimum) to 7 (maximum) according to the efficacy of the laws. A rating of 1 means the laws are nonexistent; a rating of 7 means that the laws are well developed and enforced. These annual measures of a large number of countries on identical grading scales, enable country comparisons [5]. The same unit metric will be used here to show the readiness of each one of the developed countries concerning the NEHRS legislation framework.

Percentage of GDP for Health (A4)

This aspect has not been identified in any NEHRS Plan of the five developed countries. However, it is deemed to be useful in order to present the importance and emphasis put by the various countries on the health sector. When the percentage of the GDP spending for health by a specific country is high, then this country has got the funds for the NEHRS implementation. A4 shows the amount of money that each country spends for healthcare in comparison to the whole GDP. Figures are usually presented with a percentage. For example, in 2001 Canada spent 9.5% of the Canadian GDP on health [6]. Research has shown that no country spends more than the 14% of the GDP on health. For the normalisation of this metric on a scale from 1 to 7 the very maximum found (i.e. 14%) is divided with 7 forming again 7 uniform ranges like in the case of A1-A3 (one (1) is from 0,1% to 2.0%, two (2) is from 2.1 to 4.0 etc.)

The model in graphical representation

In this section the model framework is instantiated on three countries in order to understand how it works. Since a uniform grading scale is provided for all aspects, then by ignoring non-linearity issues, one can obtain a gross figure by adding them all up. If all sub aspects/metrics of A2 are equally taken into account (as if they were actual aspects on their own) then there are eight (8) aspects in total, each ranging from 1 to 7. Thus, the maximum grade to be expected is 56. Figure 1, shows the application of the model on three countries, namely, Greece, Romania and the UK. It looks as if Greece is more prepared than Romania for the NEHRS development. Greece is graded 25 and Romania 18. Even

though Greece has larger total grade, Romania has larger grades in specific aspects such as the Legislation framework and the Broadband Internet access. However the NEHRS preparedness of Greece and Romania is far away from the NEHRS preparedness of a developed country such as the UK, which is graded with 37.

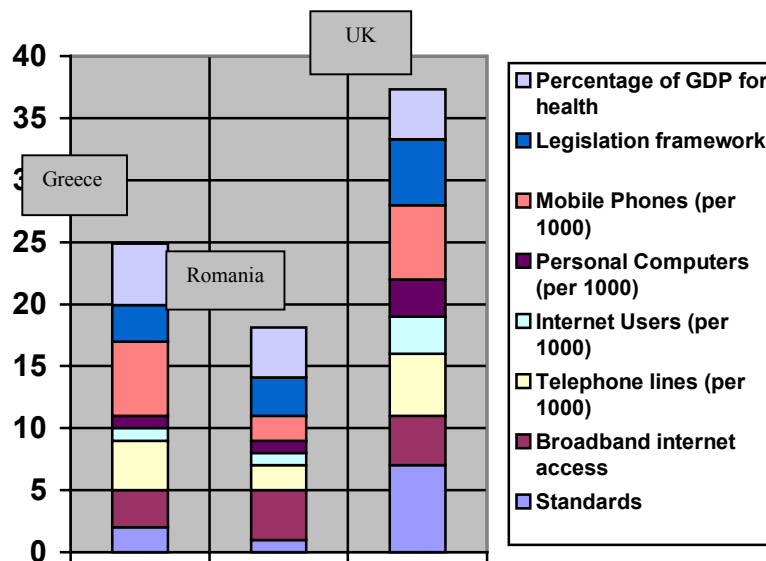


Figure 1 Application of the model framework for Greece, Romania and the UK. Each model aspect is graded on a scale from 1 to 7.

3. Delphi Method

In this section we start carrying out a historic review of the Delphi method and a description of the method's aims. Then follows an explanation of the reasons which make the Delphi method appropriate for the validation of the CM. Next we give a graphical representation of the research steps which take place during the application of the Delphi method. Finally we refer on the limitations of the method and appropriate conducts for overcoming them.

Generally the Delphi method recognizes the value of expert's opinion and experience when full scientific knowledge is lacking and accordingly, uses conformance with consensus predictions made by such experts as a criteria for judging the validity of models and theories that it is used to validate [7,8,9,10]. Clearly this is a weak test of validity, since it does not assess if the predictions derived from the theories and models, or even those of the experts, are accurate in that what is predicted will actually occur. However, in the absence of hard data or theories which would make possible a stronger test of validity, the method's rationale argues that this conformance is strongly indicative of the quality of the models or theories and hence their validity.

The first Delphi applications, outside military, were in the area of group decision process for technological, environmental and marketing forecasting. In particular, Delphi was used as a method for structuring a group communication process so that the process could be "effective in allowing a group of individuals, as a whole, to deal with a complex problem" [7]. The CM covers a wide research area making complex its validation. Apart from the CM complexity there were not any hard data or established theories that could help on its validation. A suitable validation method on the CM could have been a method predicting the validity of its results. The Delphi method was an appropriate validation method for the CM because it could forecast the validity of the model's results. The validity of the model's results correspond to the ability of the model to show the preparedness of a country to develop a NEHRS. In this way the aim of the Delphi method was to build consensus and establish a new common view on the CM where precise knowledge was missing and forecasting was necessary. Specifically we tried to validate CM predictions, on the basis that, in the absence of hard data or

established theories, predictions that were consistent with a consensus of problem domain experts were most likely to be valid. The CM was therefore considered valid if its predictions could agree with the consensus predictions of experts within the relevant problem domain of NEHRS country's preparedness.

In the above paragraph we argued the reasons that make the Delphi method appropriate validation method for CM. However, even the validation of the Delphi method has limitations because its main instrument is the opinions of experts. Experts could be wrong either because of bad application of the Delphi method or because of lack of expert's real expertise and specialization on the specific field. The conclusion is that the processes of questionnaires and selection of experts needs special care in order to apply effectively the Delphi method and achieve reliable results (opinions) [11]. Having established the general approach of the Delphi model and its appropriateness for validation of the CM, we now describe the Delphi method in detail.

3.1 An Overview of Delphi

Gordon [12] describes the Delphi method as a controlled debate. The debate is conducted by people that offer their opinion and experience on the subject, referred to as *panel of experts* [13], and is coordinated by the researcher, called the *facilitator*.

The role of the facilitator is:

- to recruit the panel of experts, and then elicit from them opinions and experience
- to prepare a series of questionnaires with which to elicit views and experience from the panel of experts. Accordingly, the questionnaires include questions formulated usually as hypotheses. Open-ended questions are suitable for the questionnaires of the Delphi method [14]. In addition, the questions follow a specific wording that enable experts to express easily their opinions and judgment. At least two rounds of questionnaire are necessary to be organized by the facilitator on this method in order to satisfy the multi-round questionnaire process basic characteristic of the Delphi method [12].
- to receive the answered questionnaires. After the feedback collection, the facilitator is processing and filtering the answers building consensus or creating new questions on newer versions of questionnaires until to build a consensus. Facilitator should keep the anonymity between the participants of the method [15].

The flowchart in figure 2 of Moeller and Shafer [2] provides a more detailed breakdown of the stages of the above Delphi method.

The above research steps of the Delphi method on figure 2 is in harmony with the suggested research steps that have been found on a number of sources during the literature research [3, 16, 17].

3.2 Limitations of the Delphi Method

When applying the Delphi method it is important to be aware of its limitations. In particular, the issues raised in the negative criticism of Delphi method found in the literature must be addressed. Sackman [18], and Armstrong [19] dispute the accuracy of the results of the Delphi method especially on complex problems. They argue that the low level reliability of expert's judgment, the poor data collection method and the difficulty of facilitator to assess the expert's opinion could influence negatively the accuracy of the results. However Goldschmidt [20] mentions that possible poor results of the Delphi method are caused because of poorly conducted Delphi projects. Poorly conducted projects can be avoided by selecting carefully the experts, and preparing questionnaires with the appropriate questions, wording and applying a pre-questionnaire for checking possible mistakes. Moreover, the facilitator should have a deep background on the research area in order to be able to manage the expert's opinions. In addition Delphi method is a qualitative research method, which is focusing more on the description and complexity of a phenomenon (issue) and not in measuring [21].

This means that the exact accuracy of the results as referred by Sackman and Armstrong is not necessary but instead of that the study of the phenomenon is the expected outcome. The expected outcome of the Delphi method is consisted of expert's opinions which reach consensus and not quantitative results where the accuracy is a critical aspect. The next sections discuss the six steps of the Delphi method which were followed in order to achieve the CM validation.

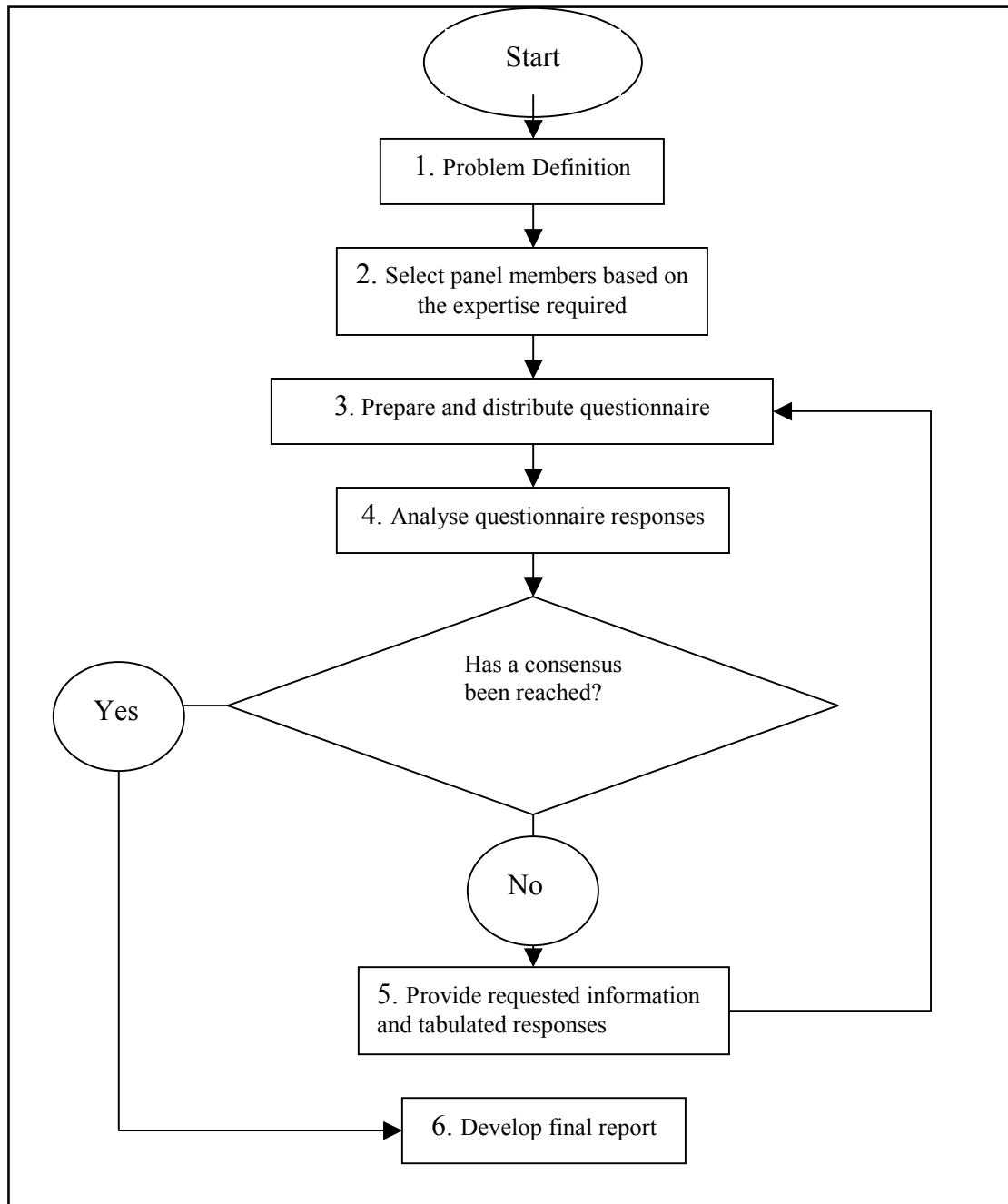


Figure 2 Flow Chart for the Delphi Method

4. Problem Definition

The problem definition of this study is close to the reason for which we prepared the CM. Specifically, analysing the NEHRS approaches of five developed countries (UK, Canada, New Zealand, USA and Australia) we led to the creation of a characterisation model including what appear to be important aspects that have an impact on the NEHRS development in these countries. The problem here is to *examine the validity of the characterisation model, and in particular to check if the included model*

aspects have strong impact on the NEHRS development and if it can really show the preparedness of a country to develop a NEHRS. It is useful to mention at this point that the circle “Start” in figure 2 includes all the literature review we made on the area of NEHRS strategies.

5. Panel of Experts

The panel of experts is the main source of data in the Delphi method and it has a similar role to that of a focus group. Focus groups are widely used in qualitative research to learn about opinion and attitudes (e.g., see [22,23]). A strength of this method of data collection is that a focus group can provide both quantitative and qualitative data, respectively by polling opinion on specific issues and using in-depth, qualitative interviews with small number of carefully selected people. However, in the Delphi method the data collection process using a panel of experts is more prescriptive and restricted than in the focus group method. Specifically, the Delphi method prescribes that data are collected from the panel using only questionnaires. This is because the experts could be dispersed all around the world and in order to reach consensus between the experts applying multi-round questionnaires on different time periods. On the other hand the focus group participants are not selected because of their expertise rather than for testing their perceptions and concerns on how they are affected by the issues being researched.

Afterwards, the panel’s composition method and its significant importance are discussed. In general, the members of the panels must have in-depth knowledge and experience on the issues where consensus is built. Adler and Ziglio [3] state that the facilitator should answer the following question after the problem definition and before the preparation of the first questionnaire: “Who are the people with expertise on the problem and where are they located?” In our case the problem has been defined in the previous section revealing the characteristics of our panel’s expertise. Specifically, the members of our panel should have been specialised on the NEHRS or the regional EHR system development.

In particular, the process of finding experts on the research area of NEHRS was difficult because a large number of experts did not want to participate in our validation phase (possibly because of time restrictions). For our model, the selection of appropriate panel members that had in-depth knowledge and experience on NEHRS programmes were: 1) academics and 2) staff which has active participation on the NEHRS programmes. These two panel participant kinds had the best possible expertise on the NEHRS area. However, the level of expertise was proved low because of the large NEHRS research area width. The problem has been mainly met on the CM aspect called ‘Standards’ which has been found difficult to be answered by the experts because there were a large variety of standards.

As regards, the location of the experts we were based on our research for literature material on the web. We found names of professors, lectures and senior lectures which had deep research background and a large number of published research articles on the NEHRS area, and exact contact details of staff participating to the NEHRS programmes in a variety of countries. Mainly we focused our efforts on finding experts from the five countries that we studied in order to prepare the CM. However we did not remain only on these five countries and we extended the search for experts to other countries as the Scandinavians, France, Germany, Holland, Malaysia and Japan. In totally we sent 53 invitations receiving 14 positive answers for participation on our research. Three of the answers were coming from staff which were working on NEHRS programmes, and the rest 11 were from the academic area. The above difference was generated because the majority of the invitations were sent to academics (40 invitations to academics and 13 to staff working on NEHRS programmes). It was easier to find on the web academics than NEHRS programmes’ staff.

Another point to stress on the panel members’ selection method is that we have not followed a standard sampling method [24, 25]. The reason is that sampling follows systematic or random participants’ selection methods from a group of people studying their perceptions and concerns on issues that affect them. On the other hand the Delphi method is not making any kind of sampling but it selects directly people which fulfill a number of expertise preconditions applying their expertise on the concerned problem which could not have a direct impact on them.

As for the number of experts that should compose of a panel, the literature was contradictory. Gordon [12] supports that Delphi study panels should consist of 15 to 35 participants. Similarly, Cabaniss [26] mentions that a panel size of 15 to 25 experts is typical for the Delphi method. Close to these two opinions is the reference of Shaw [27] which defines as small a panel having less than 10 members and large a panel having more than 30 members. Last but not least Delbecq et al. [11] indicate that the size of the experts' panel is variable and using a homogeneous group of experts then 10 to 15 participants are enough. Considering all the above sources we decided that for our case the number of experts in our panel should not be less than 10 to 15 (minimum limit) and not greater than 30 (maximum limit). Our decision derived from the average of the minimum and maximum limits of the above studies. In addition the homogeneity of our panel's expertise made unnecessary a bigger or smaller panel. The limits that we decided were reasonable because we did not need the general opinion of several people but the specific opinion of experts. It is also praiseworthy to refer the study of Murphy et al. [28] which has found few effects of panel sizes on the quality of the results comparing two same studies of six and twelve participants on each one. This conclusion helped me to justify that even if our panel has 15 or 30 experts, there are few differences on results' quality.

Another issue which considers the panel of experts is the subsequent additions to the panel of experts which are acceptable if the process starts from the beginning for the new experts [12]. It is quite common to "loose" some of the experts on each questionnaire's round. For example, it was identified in Taeyoung [29] study where Delphi methods was applied (not relevant to medical informatics) that for some of the cases, the replies of the second questionnaire round was less than 50% in comparison to the first questionnaire round replies. Certainly there is an impact on the results' quality when there are fewer responders during the second or third questionnaire rounds. The level of impact cannot be measured since none study has been found on that issue. However, it is an expected situation especially when the experts are geographically dispersed. As others have done [12, 29], we could not do something else than just accept the results of the second or third questionnaire round and make the analysis until the consensus achievement even using fewer experts. The next table 2 shows the numbers of sent and returned questionnaires on each questionnaire's version. On each questionnaire round we were "loosing" about the half experts as it was expected according to Gordon [12] and Taeyoung [29].

	1 st version	questionnaire	2 nd version	questionnaire	3 rd version	questionnaire
Number of sent questionnaires	53		14		7	
Number of received questionnaires	14		7		4	
Percentage of received questionnaires	26.4%		50%		57%	

Table 2 Number of sent and received questionnaires for each version of questionnaire.

Continuing the discussion on the panel of experts, we will now describe how we came in contact with the experts. In total, we sent 46 e-mails receiving 7 replies on our first questionnaire. We also received 7 replies during our participations on two conferences which had sessions relevant to NEHRS. These conferences were the 10th International Symposium for Health Information Management Research (iSHIMR 2005), and the 15th International conference on Control Systems and Computer Science (CSCS-15 2005). Furthermore, during the attendance of the above conferences we had very useful and progressive discussions with a number of experts on the NEHRS research area.

6. Questionnaires

In this section, we explain how we came in contact with the experts and how we designed the questionnaires. As it concerns the contact method and as prescribed by Gordon [12], initial contact with each expert started by sending invitation letters or e-mails, and asking for contribution to our study. The invitation included a description of the project, its objectives, details of how the individual

would be required to contribute to the research as a member of the panel of experts, including an estimate of the number of revisioned questionnaires to which they would be asked to respond. The invitations also assured the experts of anonymity. If the expert accepted the invitation, they were sent the first iteration of the questionnaires to fill in.

Starting our first iterations of the questionnaire, our included a welcome message similar with the one that we had in the invitation letter in order to remind panel members of the topic and the aim of the questionnaire. As for the questions' order in all the questionnaire iterations, it was random since the CM aspects had the same weight and any other question ordering was not adaptable.

Apart from the open-ended questions we have also included structured questions on each of the CM aspects in order to learn expert's quick opinion-intention for the case where the experts just agree to an aspect and they do not have to add any further comments or for the case where further opinion could be written even if the expert agrees or not. For the inclusion of these two cases the design of the questions was based on the "Tree Approach" [30]. In this approach, the questionnaire questions are organized such that the responder at the beginning give his general opinion on a structured question and if it is a necessity, further details are asked on one or more open-ended questions. More specifically, the questionnaires were including structured questions with 'YES' or 'NO' answers, supplemented by open-ended questions to allow each expert to elaborate on their thinking behind the yes/no answers.. Moreover, as recommended by Wilson [30], the relationship between a structured question and its associated open-ended question was made visually apparent by question numbering and layout conventions, whereby structured question x was supplemented by a following indented questions x.1, as illustrated in the example below (figure 3).

...Before each question there was a description on the discussed CM aspect.

2. Are there any other standards that are necessary to be included?

(Make Bold your answer or put your answer in a circle)

NO YES



2.1 If YES could you refer the name of the Standard(s):

3. Could this unit metric show the readiness for NEHRS development of a country in the area of health informatics standards?

Figure 3 Example of a questionnaire question following the "Tree Approach"

The main purpose of the structured and semi-structured (or else open-ended) questions was to establish how many of the experts, agreed with the importance of the aspects which are incorporated as unit metrics in the CM, the suitability of the unit metrics as a quantification of an aspect, and the appropriateness and usefulness of the graphical representation of the model and its aspects. As we have already mentioned the aim of the structured questions was just to learn expert's quick opinion-intention on the questions and thus, they were a source of quantitative data from which to determine if there existed consensus support for the different aspects of the CM design and its representations. On the other hand the purpose of the supplementary open-ended questions was to collect qualitative data in which the experts explain their personal stance with respect to the model and its representations. In addition, we intended that the experts would also use the supplementary questions to record their opinions on how the CM should be modified and elaborated to better model national readiness for an NEHRS.

Following good practice in questionnaire design, and as presented by Wilson [31] and Burgess [32], care was taken to ensure that questions were formulated and worded to be straightforward, concise and unambiguous. Moreover we tried to avoid double questions but contrarily we included straightforward questions [32]. In some case the questions could be called “leading questions” because they presume the knowledge on the research area of NEHRS. But this is logical because for that reason we are asking the opinion of experts on this research area. Furthermore there are questions that are called “negatives” because we were asking the experts to say if they agree or disagree. According to Hoinville [33] this could confuse the responder when he is reading the question for a first time. However considering that we were sending the questionnaire to scientists with deep background on the specific research area and having prepared them about the necessary time (approximately 30 minutes) for its completion, we believe that we overcame this possible problem. We did not have any negative comment on this and we think that it considers different cases.

In addition after the preparation of the questionnaire we ran a pilot questionnaire in order to find out possible problems on the questionnaire. This is a suggestion of Burgess [32] which we followed and we gave the questionnaire to an expert that we knew. After his answer we made a few phrasing corrections and we started to send the invitation letter first and the questionnaire secondly to the list of experts that we had already found.

The size of the questionnaire was an aspect that concerned us during the questionnaire preparation. It is well known that the responders do not even start completing the questionnaires if the questionnaire is large. In our questionnaire we had to give an understandable description for each model aspect and this was increasing the questionnaire size. We tried to reduce the size of the questionnaire and the result was 5 and a half pages. The expert needed about 30 minutes to answer the first version of the questionnaire. The second and the third versions of the questionnaires were about two pages and the experts were needing about 10 to 15 minutes.

7. Validation Results: New Version of the CM

The fourth stage includes the experts’ answers elaboration which is followed by a loop, and if consensus between the answers is reached then the CM aspect is accepted and added on the final version of the CM (sixth step) or if no consensus is achieved, the fifth step is taking place where a revision on the questionnaire is taking place in order to achieve consensus. The methodology on which we were based for engineering the questionnaire answers and achieve the consensus was the following: firstly we measured the positive and negative answers for each aspect examining the answers on the structured questions; secondly we elaborated the answers on the open-ended questions; thirdly if there was one or more strong arguments on the open-ended questions then the result of the open-ended questions (accept the aspect or modify it or deleted it) was prevailing even if the structured questions were supporting the opposite result. The prevailing of the open-ended questions is caused because the greater importance and weight of the justified expert opinions and also because the structured questions were there for helping us in achieving the quick opinion-intention or in case the evidence of the open-ended questions was not enough for concluding to a result. However, for brevity’s sake the experts’ replies and questionnaires’ revisions which took place in all the questionnaire rounds are not referred in the paper. We only mention the contradictive issues which were considered during the questionnaire process and the validation results.

The issues which were raised during the Delphi method application are the following:

- Inclusion of a number of extra standards.
- New categories of standards, changes on the existent categories and transposition of standards from one category to another have been suggested in order to avoid standards overlapping.
- Fundamental changes on the aspect “Health Network Architecture” as some of the subaspects do not correspond to the aspect name.
- Elimination of some subaspects to avoid proper weighting problems with the other aspects.

- Inclusion of two extra aspects such as the computer literacy of the health personnel and the health network infrastructure (network's link capacity).
- No consensus has been reached on whether the aspect “%GDP for health” should be considered in the model or not because for some of the experts this aspect does not represent actual NEHRS expenditure.
- No consensus has been reached until now on whether the unit metric “Legislation framework” represents the readiness of a country on legislation for a NEHRS or not.
- It has been suggested to separate the NEHRS readiness of a country on categories: 1) Governmental (Laws, ICT infrastructure), 2) Health Sector (Standards, Computer Literacy of health personnel, Broadband Network Infrastructure) and 3) Citizen (computer literacy, available PCs and Internet links).

7.1 Example of the new Characterisation Model

In this section an example of the new CM applied on Australian situation is presented. In particular we give an overall representation of the CM aspects and its graphical representation. The analysis of the questionnaires concluded to a number of important changes on the characterisation model of section 2. One critical modification that was suggested by the experts was the separation of the country readiness in different categories according to the sectors that NEHRS readiness aspects correspond. The governmental (national) readiness, the health sector readiness and the citizen readiness were the three proposed categories of readiness. In all the NEHRS aspect, unit metrics have been adapted from 1 to 7 where the 1 is the minimum and 7 is the maximum. An overall description of the model aspects and its unit metrics including the sources or the tools, which help in the calculations, is the following:

The Governmental (regional) Readiness shows mainly the ICT infrastructure from technological side but also from the legislation side. The availability of fast Internet access and the number of telephone lines across a country indicate the national ICT infrastructure which is important for the NEHRS operation. The legislation infrastructure for ICT is a governmental responsibility to be prepared and a very critical aspect for the normal NEHRS operation. This aspect has been extensively discussed in the previous paragraphs. Altogether the aspects of this category are the following:

1. Governmental (Regional) Readiness

1.1) Broadband Internet Access Availability (1 to 7) – (source World Bank).

1.2) The number of telephone lines per 1000 people (Normalised from 1 to 7). In order to achieve normalisation of the metric the 1000 is uniformly divided into 7 equal ranges, in order to keep the same normalisation with other metrics of the model (i.e. the unit metric ranges from one (1) (1 - 142,9) to seven (7) (857.4-1000).

1.3) Laws related to ICT use (1 to 7) – (source World Bank).

The Health Sector Readiness indicates the preparedness of the health sector of a country on different but equally important aspects. The first aspect is the standards which support mainly the data security and availability (interoperability). Furthermore the computer literacy of the health personnel is calculated in order to evidence the readiness of the health personnel to accept the NEHRS (acceptability). In addition the network infrastructure that has the health centres is estimated. This aspect intends to show the readiness of the health centres to adapt different NEHRS technologies. Altogether the aspects of this category are the following:

2. Health Sector Readiness

2.1) Health Informatics Standards (1 to 7) – (Table of Standards).

2.2) Computer literacy of the health personnel (1 to 7) – (questionnaire).

2.3) Broadband Internet (Network) Infrastructure in the health centres (1 to 7) (Governmental data).

The third readiness category which is called “Citizen Readiness” covers the computer literacy of the citizens aspect. Moreover this aspect intends to define the number of citizens that are familiar with mobile devices and have the ability to have access to their records from remote locations. Altogether the aspects of this category are the following:

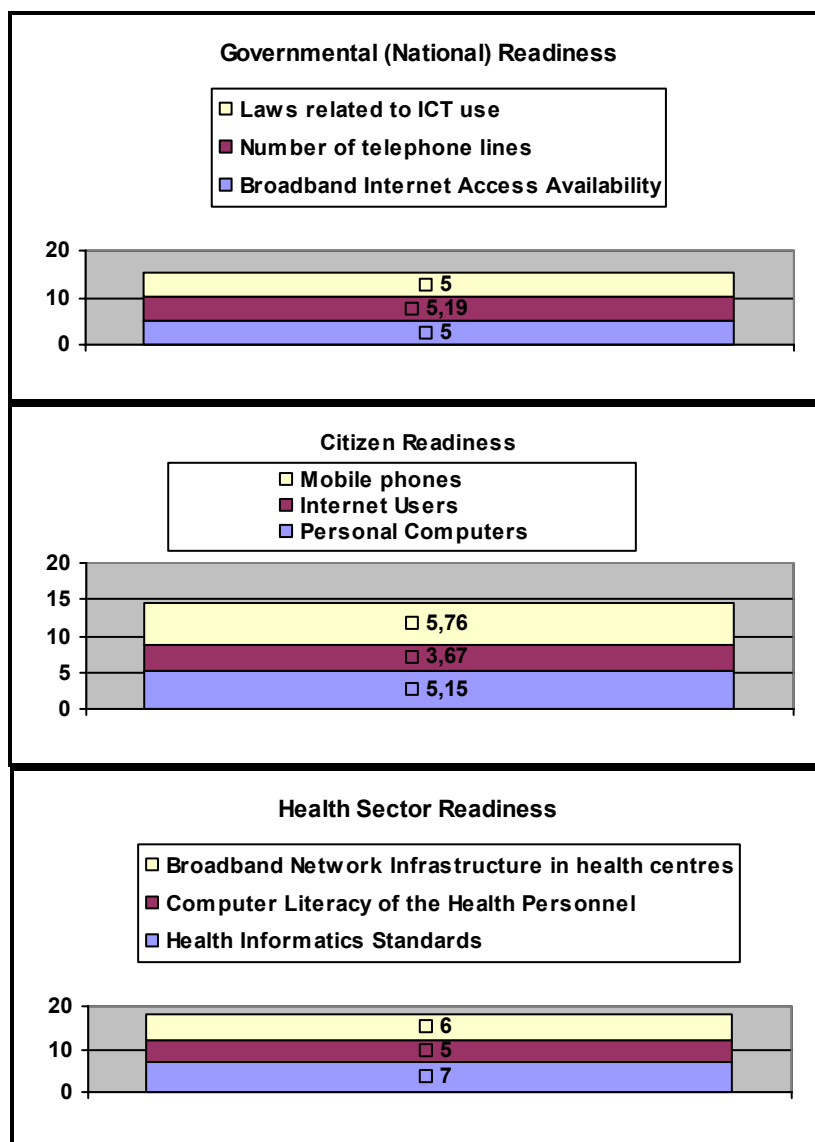


Figure 4 Graphical representation of the Characterisation Model

3. Citizen readiness

3.1) Personal computers per 1000 people (Normalised from 1 to 7). In order to achieve normalisation of the metric the 1000 is uniformly divided into 7 equal ranges, in order to keep the same normalisation with other metrics of the model (i.e. unit metric ranges from one (1) (1 - 142,9) to seven (7) (857.4-1000)).

3.2) Internet Users (thousands) (Normalised from 1 to 7). The World Bank gives the number of Internet Users in thousands where knowing the total population of a country it can be calculated the number of internet users per 1000 people $((\text{Number of Internet Users}) \times 1000) / \text{Total population}$. In order to achieve normalisation of the metric the 1000 is uniformly divided into 7 equal ranges, in order to keep the same normalisation with other metrics of the model (i.e. the unit metric ranges from one (1) (1 - 142,9) to seven (7) (857.4-1000)).

3.3) Mobile Phones per 1000 people (Normalised from 1 to 7). In order to achieve normalisation of the metric the 1000 is uniformly divided into 7 equal ranges, in order to keep the same normalisation with other metrics of the model (i.e. the unit metric ranges from one (1) (1 - 142,9) to seven (7) (857.4-1000).

The graphical representation of the characterisation model is shown on figure 4.

8. Summary and Conclusions

This paper has described the CM and the process by which it has been validated. In the first part of the paper we presented the CM which were created using the literature and in the second part we described the Delphi validation method that we have adapted on our model. In the third part, the results of applying that process were presented together with the modified CM after the Delphi validation. In this last part of the paper, we are summarizing the paper referring also a number of conclusions on the validity of the CM.

The first version of the CM was derived through review and analysis of secondary data in the literature. It was therefore necessary to establish the validity of the model. However, this was problematic due to the lack of hard data or established theory relating to NEHRS. The lack of even one complete NEHRS restricted our validation options. Consequently, it was decided to test the validity of the CM by establishing the extent to which its properties and behaviour are consistent with the consensus opinions of experts on NEHRS, using the Delphi method.

As regards the Delphi validation method implementation on the CM, the experts were asked to proffer their opinion on the construction of the CM and its consequential capability to estimate the readiness of a country to develop a NEHRS. More specifically, the intention was to solicit the expert's opinions about each model aspect, the unit metric corresponding to each aspect, and the graphical model representation. Furthermore we wanted to solicit their suggestions for possible elaboration and enhancement of the model, for example through addition, removal or redefinition of its component aspects. As prescribed in the method (ref), opinions of the experts were solicited using questionnaires. The questionnaires were designed with both structured and semi-structured questions. After three rounds of questionnaires we ended up to a new CM much different than the one that has been presented in section 2.

As for the conclusions that were raised during the CM validation these are: 1) The Delphi method is appropriate for the validation of such a model since a number of deficiencies have been identified and covered. After the Delphi method application the CM has improved the quality of the results that it provides. 2) The most critical and difficult part of the Delphi method was the communication with the experts. Even if the Delphi method was appropriate for our model, the experts' participation in all the questionnaire rounds was proved a serious problem of the Delphi method. The general conclusion is that a panel of experts which is geographically dispersed could cause problems on the validation by avoiding to participate in all the questionnaire rounds.

Finishing this paper, it worths to mention that the CM is only a piece of the models and methodology that is suggested from this research as evidence based know-how for developing NEHRS. Any country can estimate its NEHRS readiness identifying specific deficiencies on its infrastructure.

References

- 1 Lee, A.S., "Electronic Mail as a Medium for Rich Communication: An Empirical Investigation Using Hermeneutic Interpretation," MIS Quarterly, Volume 18, Number 2, 1994: 143-157.
- 2 Moeller, G.H. & Shafer, E.L., Ch. 39 in Ritchie and Goeldner, and Taylor, R.E. & Judd, L.L. (1994). "Delphi Forecasting" in Witt, S.F. & Moutinho, L. Tourism Marketing and Management Handbook. London: Prentice Hall.
- 3 Adler, M., & Ziglio, E. (1996). Gazing into the oracle. Jessica Kingsley Publishers, Bristol.

- 4 Orfanidis L., Eaglestone B., and Bamidis P. (2005) 'A Model Framework to Characterise the NEHRS Preparedness: a South-East European perspective', In the Proceedings of the 15th International Conference on Control Systems and Computer Science, Bucharest, Romania.
- 5 World Bank (2003) 'ICT at a glance', www.worldbank.org March 2004.
- 6 World Health Organization (2003) 'Canada', www.who.int/country/can/en/ March 2004
- 7 Linstone, H. A. & Turoff, M. (1975). "The delphi method: Techniques and applications". Addison-Wesley Publishing Company, Reading. <http://www.is.njit.edu/pubs/delphibook/> March 2006.
- 8 Keeney, S., Hasson, F.& McKenna, H. (2001). A critical review of the Delphi technique as a research methodology for nursing. *International Journal of Nursing Studies*, 38: 195-200.
- 9 Rowe, G & Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting*, 15: 353-375.
- 10 Snyder-Halpern, R. (2000). Comparison of Mailed vs. Internet Applications of the Delphi Technique in Clinical Informatics Research. AMIA Annual symposium. <http://www.amia.org/pubs/symposia/D200120.PDF> May 2005
- 11 Delbecq, A., Van de Ven A., & Gustafson D. (1975). Group techniques for program planning: a guide to nominal group and Delphi processes. Glenview, Illinois: Scott Foresman.
- 12 Gordon Th. (1994) 'The Delphi Method', AC/UNU Millennium Project http://www.futurovenezuela.org/_curso/5-delphi.pdf April 2006
- 13 Wikipedia (2005) "Delphi method". Wikipedia: the free encyclopedia, http://en.wikipedia.org/wiki/Delphi_method March 2006
- 14 Kurth-Schai, R., Poolpatarchewin, C. & Pitiyanuwat, S. (2000). Using the Delphi Cross-culturally: toward the development of policy. In John Cogan & Ray Derricott (Eds.), *Citizenship for the 21st Century* (pp.93-108). London: Kogan Page Ltd.
- 15 Turoff, M., & Hiltz, S. (1996). Computer-based Delphi Processes. In M. Adler & E. Ziglio (Eds.), *Gazing into the oracle. The Delphi method and its application to social policy and public health* (pp56-85.). London: Jessica Kingsley Publishers.
- 16 Fowles, J. (1978) 'Handbook of futures research', Greenwood Press: Connecticut.
- 17 Hanson, W.H., and Ramani, N. (1988). Technology forecasting: a hydroelectric company experience. *Technology Management Publication, Harvard Business School, Press*.1(3): 266-270.
- 18 Sackman H. (1975) 'Delphi Critique – Expert opinion, forecasting, and group process', Lexington Books, Toronto.
- 19 Armstong J. (1999) 'Introduction to paper and commentaries on the Delphi techniques', *International Journal of Forecasting*, Vol. 15: 351-352.
- 20 Goldschmidt P. (1975) 'Scientific inquiry or political critiques?', *Technological forecasting and social change*, Vol. 7: 195-213.
- 21 Arksey H. and Knight, P. (1999) 'Interviewing for Social Scientists', Sage Publication, London: 1-20.
- 22 McNamara C. (1999) 'Basics of Conducting Focus Groups', *Guidelines for Conducting Focus Groups*, http://www.managementhelp.org/grp_skll/focusgrp/focusgrp.htm January 2006
- 23 Luntz, F (1994) 'Voices of Victory, Part I: Focus Group Research in American Politics', *The Polling Report*, <http://www.pollingreport.com/focus.htm> January 2006
- 24 Seale Cl. And Gobo G. and Gubrium J. and Silverman D. (2004) 'Qualitative Research Practice', Sage Publications, London:68-71.
- 25 Blaxter L. and Hughes Ch. And Tight M. (2001) 'How to Research', Open University Press, Buckingham: 160-177.
- 26 Cabaniss K. (2001) 'Counseling and computer technology in the new millennium – An Internet Delphi study', Doctoral Dissertation, Virginia Polytechnic Institute and State University, Department Counselor Education.
- 27 Shaw Z. (1981) 'Group dynamics – The psychology of small group behavior', McGraw-Hill Book Company, Toronto.
- 28 Murphy M., Black N., Lamping D., McKee C., Sanderson C., Askham J., Martineau T. (1998) 'Consensus development methods and their use in clinical guideline development', *Health technology assessment*, Vol. 2(3).

- 29 Taeyoung Shin, Delphi study at the multi-country level: Gains and Limitations, In the Proceeding of the International Conference on Technology Foresight 2000, Tokyo, Japan.
- 30 Wilson T. (2002) 'Designing questionnaires and interview schedules', In: On-line course on designing questionnaires and interview schedules, <http://www.hb.se/bhs/nyutb/kurswebb/c-kurser/applirm/gammalt/qdes1.html> January 2006
- 31 Wilson T. (1985) 'Questionnaire design in the context of information research', In: M. Brenner, J. Brown and D. Canter, eds, *The research interview: uses and approaches*, p: 65-77, <http://informationnr.net/tdw/publ/papers/1985qdesign.html> January 2006
- 32 Burgess T. (2001) 'A general introduction to the design of questionnaires for survey research', Information systems services, University of Leeds, <http://www.leeds.ac.uk/iss/documentation/top/top2.pdf> January 2006
- 33 Hoinville G. and Jowell R. and associates (1978) 'Survey research practice', Heinemann Educational Books, London, In: Wilson T. (2002) 'Designing questionnaires and interview schedules', In: On-line course on designing questionnaires and interview schedules, <http://www.hb.se/bhs/nyutb/kurswebb/c-kurser/applirm/gammalt/qdes1.html> January 2006.

Investigation of Emergent Phenomena within Complex Systems

Ognen Paunovski¹, George Eleftherakis², Tony Cowling³

¹*SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, ogpaunovski@seerc.info*

²*City College, Dept. of Computer Science, 13 Tsimiski Str, 54624 Thessaloniki, Greece, eleftherakis@city.academic.gr*

³*The University of Sheffield, Dept. of Computer Science, Western Bank, Sheffield, S10 2TN, UK, ajc@dc.shef.ac.uk*

In recent years the phenomenon of emergence as part of complex systems has significantly gained in importance. This is partially due to the problems faced by information and communication systems, as well as the potential for advancing the engineering of adaptive self-organizing systems. The research area appeared as a natural response to the failure of traditional reductionistic methods when dealing with dynamic organization and behaviour in complex systems at runtime. In order to address this problem agent-oriented modelling and simulation has been used as the main tool for exploring emergence within the scope of computer science. The initial idea is to model simple local rules for individual agents that would be able to generate emergent organization and behaviour at the system level. Then by means of simulation and analysis to gain insight into the emergent runtime phenomena. This paper overviews some of the major issues concerning emergence in complex systems. This includes the definition of complex system and emergence, types of emergence as well as multi-agent based modelling of emergent behaviour.

Keywords

Complex systems, emergence, types of emergence, multiagent modelling.

1. Introduction

The world today relies on globalization accompanied with increasing interconnection in order to devise cheap, simplified working processes that would increase efficiency. The major support for this trend is based on the constant innovation in the Information and Communication Technology (ICT). Examples of ICT support of interconnection can be found in all aspects of everyday life, like the increase in the mobility with mobile telephony and wireless networks, cultural exchange through the Internet, online governmental services, economic interconnection (from manufacturing to logistics, retail to finance) and so on.

However as ICT systems expand, they form largely unplanned and unregulated interconnections, thus increasing their complexity, causing problems in the design, management, maintenance and decommission of such systems [1]. The classical strategies based on reductionistic methods cannot deal with this complexity increase and begin to fail. The reason can be found in the inability to predict and control runtime emergent properties of the system which are based on stochastic interaction between elementary components and/or external entities. The problem diminishes the ability to give functional and performance guaranties for open ICT systems with high complexity. On the other hand controllability of emergent behaviour can be very useful in systems engineering. Possible advancements in the development of self-organizing adaptive systems could revolutionize many aspects of software engineering.

Therefore in recent years significant research efforts have been directed towards understanding emergent phenomena within complex systems. In many cases agent-oriented computer modelling and simulation has been used as tool for exploring emergent behaviour. The idea is to use simple reactive agents in order to model complex system exhibiting emergent phenomena. The developed models can then be used in order to gain a deeper understanding of the phenomenon. In this context one of the essential aspirations of such attempts is to capture the relation between local and global behaviour, also called the Micro-Macro Link (MML) relation. Many scholars believe that understanding this link is essential for understanding the nature of the emergent phenomena.

This paper deals with the theoretical background of some of the elementary concepts involved in modelling and understanding emergent phenomena within complex systems. The section that follows examines the concepts of complexity and complex system. Section 3 focuses on emergence by providing a working definition as well as details of different types of the phenomenon. In addition section 4 focuses on the relationship between multiagent and complex systems, it also explores the possibility of modelling emergent behaviour with multiagent system. The final section summarizes some of the major points and future work.

2. Complexity and complex systems

In broad terms Complexity theory is a multidisciplinary field of research dealing with Complex Systems. There are many examples of such systems existing in nature. Cells, embryos, nervous systems, immune systems, ant colonies, etc; can all be viewed as complex systems. In the human world, cultural and social systems like political parties, scientific communities etc are also complex systems. Nevertheless these examples do not provide insight into what complex system is. An attempt at understanding and defining a complex system must start with a definition of complexity. A general view of complexity suggests that it is a property of a specific entity (presumably a system) that could be quantified. Therefore it seems logical to allocate the role of a metric to the concept of complexity. In fact there are many ways to measure complexity. For example computational complexity [2] denotes the complexity of a sequence of symbols by describing the system as finite state machine that produces this sequence. On the other hand biologists distinguish between structural and functional complexities when describing an organism, there are also many other complexity metrics used in specific domains (see [3] for more information). While each measure is useful in its own domain, there is a fundamental problem in measuring complexity. This is the inability to express complexity in a scale that will be universally accepted in all domains. In other words there are no means for measuring complexity applicable in all possible cases. This means that there is no common scale to measure and/or compare the complexity of a piece of text and biological organism. Given the nature of the problem, it seems highly unlikely that it will be solved in the near future. Consequently most of the researchers in the area of complex systems, use descriptive approaches to define complex systems. Nevertheless there is no agreement in the scientific community on a single definition. As a result there is a variety of definitions provided by different authors. For example Badii and Polity in [4] used the definition from Webster dictionary where a complex system is defined as:

“A complex object is an arrangement of parts so intricate as to be hard to understand or deal with”

On the other hand Holland’s view expressed in [5] is that:

“A complex adaptive system is complex, self-similar collection of interacting adaptive agents”

In fact the whole situation is best described by Simon in [6] where he says:

“I shall not undertake a formal definition of complex systems. Roughly by a complex system I mean one made up of a large number of parts that interact in a non-simple way”

If the definitions provided in literature are closely examined, the conclusion is that most of them revolve around two major points. First complex systems are made up of many interacting parts and second they are “not easy to be understood”. It should be noted here that the second point, referring to

the concept of “not easy to be understood”, means that the system is highly complicated (or complex). Therefore one might argue that this is a subjective measure in the sense that something which is easy to understand for one person might be complicated for another. Nevertheless as complexity increases such claims become invalid. Thus it is absurd to claim that the global ecosystem of this planet is a system which is easy to understand and predict at very fine details. In this context one of the fundamental problems in predicting and understanding complex systems is the phenomena described as emergence. The next section examines this concept more closely.

3. Emergence in complex systems

Emergence is one of the most important and at the same time one of the least understood characteristics of complex systems. The sections that follow discuss the main concept behind emergence, the different types of emergence as well as its connection with self-organization and evolution. In addition the possibility of utilizing emergence in software engineering is examined in the final section.

3.1 Defining Emergence

The term “emergence” (from Latin “emergere”) means “to become apparent”, “to turn up”, “present itself”, “to appear” (Oxford dictionary). The whole idea behind emergence was popularized in 1972 by Anderson in [7] where he elaborated how global features may arise (emerge) as a property of a system and at the same time to be novel as to the constitutive components of the system. In other words Anderson’s definition of emergence is the manifestation of emergent properties in terms of high level structures and organization that cannot be identified in the sub-components of the system. This basic concept behind emergence is generally used to describe a system where global phenomenon arises from the local interactions between the micro level components of that system. For example one H₂O molecule is not a fluid, but millions of H₂O molecules at room temperature have the property of liquidity. Another example would be the concept of consciousness (in the human brain) which is a result of collective interplay of millions of neurons, where a single neuron can not give rise to such phenomenon. In the natural world global pheromone paths arise from pheromone dropping ants following local paths.

The phenomenon of emergence is studied from different aspects in different sciences, leading towards problems in constructing a generally accepted scientific definition. In fact different research fields have adopted different views useful to their own discipline. In order to continue the discussion on emergence, there is a need for a working definition in the framework of computer science. After extensive literature, this paper adopts the definition provided by Wolf and Holvoet in [8]:

“A system exhibits emergences 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.”

In this context “coherent emergents” refers to orderly (logically or aesthetically) consistent results of the process of emergence which could be properties, behaviour, structure, pattern, etc. The Macro-level denotes the global view of a system (e.g. a Multi-Agent System (MAS) as a whole), while the micro-level refers to a view of a system as a set of basic individual components (e.g. set of agents). However given the variety of the emergent phenomena, a single definition is not sufficient in order to provide a comprehensive understanding. Simply put we need to be able to distinguish between the different emergent phenomena.

3.2 Types of emergence

The ability to distinguish between the varieties of emergent phenomena is essential in further investigation of the subject. Nevertheless, like the definition of emergence, there is no general agreement on how to differentiate and classify emergent phenomena. The most widely used

classification is based on the notions of strong and weak emergence [9, 10]. In this context strong emergence is defined as theoretically irreducible to the properties of the fundamental elements of the system. On the other hand weak emergence is an unexpected phenomenon given the properties and interaction of the basic elements. In addition Bedau [10] also uses nominal emergence in order to describe a global property of the system that cannot be the property of the fundamental elements. However this classification based on the philosophical study of the phenomenon is insufficient to resolve the classification problem for applied sciences like physics, biology, computer science etc.

In order to achieve more unambiguous definition of the types of emergence it is crucial to devise appropriate classification criteria. Heylighen in [11] proposed the following criteria as a basis for the classification of emergence.

- Amount of variety of possible states - refers to the change in the number of states of the system as a result of the emergence.
- Internality/Externality - refers to the origin of the variation and selection mechanisms influencing the emergent phenomena.
- Number of levels - refers to the number of levels of emergence (the usual distinction is between single or multilevel cause and effect relations).
- Contingency of constraint - refers to the required influence by specific processes in order to form and maintain emergents.

It should be noted that the criteria presented here are not completely independent, for example a multilevel emergence process exhibits much more variety than single level emergence and imposes more constraints on the system. Consequently Fromm [12] argues that it is possible to reduce the classification criteria and distinguish between the different types according to the causality structure of emergence. Thus he proposes a classification criterion which is based on the so called cause-effect relationship. In this context the process of emergence can be viewed as an effect where the cause is not immediately apparent. This approach actually builds upon the classification for cellular automata proposed by Wolfram [13]. The resulting classification based on this idea is the following:

- Type I – Simple/Nominal emergence. This type of emergence is very common and contains no top-down feedback, but only bottom-up feed-forward relations. Usually this type of emergence is exhibited by closed systems with passive entities which don't change their properties or behaviours. Although a system of this type can be highly complex, it is fully predictable.
 - Type IA – Simple intentional. The emergent property of this type is actually designed, in the sense that there is a planned and controlled interaction of the elementary components in order to produce the global pattern. For example an intentional design of a machine (clock, TV, software application), the semantics of a sentence (as emergent property of arrangement of words), etc.
 - Type IB – Simple unintentional. This subtype of emergence is common in systems with a large number of loosely coupled identical elements. The emergence in these systems is a result of interaction between particles which influences statistical quantities like the thermodynamic properties of pressure, volume and temperature. Examples of this type of emergence could be a wave front in water, avalanches, cascades and so on.
- Type II - Weak emergence. Emergence with single top-down feedback which could be negative (type IIA) or positive (type IIB). Usually exhibited by open systems with active entities. Emergence of this type is predictable in theory but not always in fine details.
 - Type IIA - Weak stable. This is a stable form of emergence with bottom-up cause and top-down negative feedback acting as a constraint on the low level components. It is one of the most studied types of emergence because it has great potential for commercial application. Examples of this kind of emergence include flocking of birds, ant colonies, different forms of self-organization (e.g. WWW, Wikipedia), etc.

- Type IIB - Weak instable. This form of emergence is usually referred as negative emergence due to the fact that is based on imitation (self-amplification). Simply put the emergents rely on reinforcement through positive feedback that leads to violation of equilibrium due to exponential growth. Examples of this type of emergence are: crashes and bubbles in the stock market, explosions of social unrest, buzz in the news, celebrity effect, etc.
- Type III - Multiple emergence. Emergence with multiple feedbacks, both positive and negative, common for open systems with high complexity. These systems are usually composed out of components with fluctuating roles. Consequently the systems exhibiting this kind of emergence are highly dynamic comprised of multiple levels. This type of emergence is not predictable due to its chaotic nature.
 - Type IIIA - Multiple feedback. This type of emergence is a combination of IIA and IIB, usually based on short ranged activation (positive feedback) leading to long term inhibition (negative feedback). This type of emergence is strongly related to activator-inhibitor systems. Examples of this kind of emergence can be found in pattern formation in biological entities (leopards, jaguars), stock market rush, prisoners dilemma and so forth.
 - Type IIIB - Adaptive emergence. This type of emergence is often associated with evolutionary transitions and the appearance of complexity. Thus it represents a jump in complexity in terms of bridging a fitness barrier (by catastrophe, exaptation and tunnelling). Examples include evolution of ecosystems, sudden scientific or mental revolutions, extinctions and mass catastrophes, and so on.
- Type IV - Strong emergence. Strong emergence has the highest complexity in this classification. It basically denotes the emergence of structures on higher levels of organizational complexity with very large jumps in complexity and/or major evolutionary transitions. Due to the complexity of this type of emergence it is not predictable even in theory. The most obvious example of strong emergence is life as an emergent property of genes, the genetic code, nucleic amino acids. Similarly the social concept of culture can be viewed as a strong emergent of memes, language and writing.

Type		Name	Roles	Frequency	System	Feedback
I	IA	Simple	Fixed	Abundant	Closed, with passive entities	No feedback
	IB					Scale preserving (peer to peer) feedback
II	II	Weak	flexible	Frequent	Open, active entities	Scale crossing (top down) feedback, positive and negative
III	IIIA	Multiple	fluctuating	Common	Open, multiple levels	Scale crossing (top down) feedback, positive and negative
	IIIB					Multiple feedbacks in a system
IV	IV	Strong	New world of roles	Rare	New or many systems	All above including feedback between different systems

Table 1: Summary of the major properties of the different types of emergence, taken from [8].

As is visible from Table 1, the complexity of the emergent phenomena increases as we move from simple towards strong emergence. At the same time as the complexity increases the system becomes more dynamic and thus less predictable. Nevertheless, as the emergent phenomena become more complicated the number of systems exhibiting them decreases.

3.3 Emergence, Evolution and Self-organization

Evolution and natural selection are the main driving forces of complexity increase in nature. As such they can be a cause for emergence. Usually evolution is closely associated with the strong forms of emergence, more specifically type IIIB and type IV. In the case of type IV the emergent behaviour is caused by a sudden jump in complexity (see [14] for more information). The jump is a major evolutionary transition that has created completely new set of roles or systems. On the other hand type IIIB which is related to adaptation and learning is more closely related to natural selection.

Another interesting concept that is often associated with emergence is self-organization. On many occasions self-organization is used as a synonym for emergence. This creates confusion since each concept describes a completely different characteristic of the system and could exist in isolation from the other. Self-organization is usually used to describe a system that appears to organize itself without external direction, manipulation or control. A system can exhibit emergence without self-organization and vice versa. Nonetheless there are several important similarities between the two phenomena. Both of them are dynamic processes which arise over time [15]. Also both demonstrate robustness. In the case of emergence the robustness is viewed as the ability to withstand failure of a single component without the failure of the emergent property. On the other hand the robustness in the case of self-organization is a product of adaptability to changes by increasing the order within the system. These properties are complementary and therefore are very useful in combination. Therefore it comes as no surprise that engineering self-organizing emergent systems is a major issue for software engineers.

In order to develop a self-organizing emergent system there are two main approaches. The first approach considers self-organization to be the cause for emergent functionality. In other words the emergent pattern at the macro level can arise as result of a designed self-organizing process. There are several basic principles like feedback loops, stochastic relationships, balance, and so forth, which are used in order to construct self-organizing emergent systems in this approach. Nevertheless in order to achieve the desired results the elementary components usually need to be proactive with well designed interaction patterns. The second approach follows a different direction. It starts with the emergent property as a force that leads towards organization of the system. In this approach the self-organization is a product of relatively simple elements and the interaction between them. This design significantly reduces the complexity of the micro level components and by this rationale it is easier to be developed. However due to the reactive design of the components and the unpredictability of the interaction it is difficult to test and give guarantees about the runtime properties (functionality) of the system.

3.4 Utilization of Emergence in Software systems

Emergence is a crucial characteristic of many complex systems existing in nature as well as a property of many social systems. The versatility of the phenomenon shows great potential for problem solving in many domains. Software engineering is one of the domains that can significantly benefit from the utilization of emergent phenomena. In this context the emergent properties exhibited by biological systems can present useful solutions to a variety of problems faced by ICT systems today. For example social-insects such as ant colonies or bee hives [16, 17] demonstrate survivability, adaptability and persistence within an open interaction environment. These kinds of characteristics can be very useful as parts of software systems that operate in an open dynamic environment like the Internet. Although many concepts from biology were introduced into computer science in areas like networks, security, pervasive computing and so on, most of the attempts were aimed at isolated problems within specific settings. In order to utilize the full potential there is a need for a general development framework for engineering emergent self-organizing applications.

However there is a major obstacle, self-organizing emergent design can only be accepted in the industry if one can give guarantees about the macroscopic behaviour and/or macroscopic organization. In other words the designer of the system must be able to somehow prove (formally or informally) that the system works as intended. Nevertheless given the unpredictable nature of emergence and the fact that (at the moment) there is no way to capture it with corresponding models, it is practically infeasible to formally verify or prove the macroscopic effect of emergent phenomena. Moreover Wegener in [18]

shows that all possible behaviours of a large scale open interaction model cannot be verified, thus proving that is not merely difficult but impossible to formally verify such system. Therefore the only viable solution is to use empirical approaches in order to verify the emergent macroscopic behaviour.

4. Modelling emergence with multi-agent systems

Many researchers agree that Multi-Agent Systems are the best tool currently available for exploring emergence within complex systems. The sections that follow examine some of the issues involved in modelling emergent behaviour as well as the essential properties of multi-agent models exhibiting emergence. In addition the last section contains a brief discussion on the micro-macro link problem.

4.1 Problems in modelling emergent behaviour

Despite the problems in defining and researching emergence discussed previously, the phenomenon itself poses a great deal of difficulty. First of all emergence exists in the eye of the observer, which means that identification of a specific macroscopic pattern or behaviour as emergent is subjective. In other words a specific pattern could be described as emergent by someone while for others it could be just some random noise. So the most obvious question is what emergence says about the system. Thus in most cases it is very difficult to comprehend the character of the emergent functionality. This leads to almost no success in developing methodologies for predicting its behaviour. The fundamental problem in predicting emergence is the inability to capture it with any model of the system. This is due to the fact that emergent behaviour relies on stochastic runtime interactions between the elementary components of the system. Moreover it is invisible to be analyzed using bottom-up or top down approaches making identification of interaction patterns (which presumably are responsible for the emergents) extremely difficult. In addition some specific types of strong or multiple feedback emergents are not predictable even in theory, which makes the idea of understanding and controlling these types of emergence very distant.

Another issue that was partially discussed in previous sections is the correctness of a system exhibiting emergent behaviour. Although formal methods can be used in order to ensure that the model of an individual agent is “correct”, they lack the power to deal with a dynamic multi-agent model of a complex system. This inadequacy is clearly presented by Edmonds and Bryson in [19] where they identify the limitation of formal methods when dealing with complex systems. Furthermore there is no a suitable set of elementary rules and corresponding calculus (code) to describe an emergent phenomenon.

In essence all these intrinsic obstacles can be attributed to a single problem. This is our inability to define, quantify and understand the nature of the phenomena. So the question that remains is what we can do to improve our understanding of emergence. An initial idea is to explore emergence through the modelling of complex systems using multi-agent techniques.

4.2 Multi-agent models with emergent properties

Many researchers in the field of complex systems agree that the initial approach in understanding emergence should be done through multi-agent systems. There are several reasons that correlate complex systems with the concepts behind MAS. First there is a natural correspondence between the structure of a complex system and a multi-agent system. Both of them rely on many individual components in order to achieve their goals. Furthermore each agent in MAS is autonomous and it is capable of interacting in a stochastic manner with other components/agents. Moreover there is no limitation on the interaction scenarios, which means that an agent is able to communicate indirectly on multiple levels by modifying the local environment. This is one of the major approaches used for communication in natural complex systems. Also MAS imposes no limitation on the complexity level which could be achieved. Thus a multi-agent system of specific complexity can be viewed as a complex system.

Although according to what was discussed above MAS is a natural tool for modelling complex systems, it does not mean that any multiagent system is complex or exhibits emergent behaviour by default. Nevertheless these features are more common in MAS compared to more traditional approaches like Object Oriented Programming (OOP). Fromm in [20] identified several properties common for MAS exhibiting emergence:

- Agent mobility or visible states for fixed systems – e.g. spatial repositioning for mobile agents.
- Ability to influence the environment – e.g. chemotaxis, self-replication or other approaches for modifying the environment.
- Ability to distinguish between groups and individuals – e.g. flocking of birds as a model composed of individual agents and groups (flocks).

Another issue that has been considered is the complexity of the individual components. Usually in complex systems the elementary components follow a relatively simple and straightforward design. By this analogy the components of a complex system more closely resemble reactive (simple design with few operational rules) than proactive agents (agents with logical representation of the world). Furthermore the impact of the emergents within proactive MAS is significantly reduced due to the context dependence of the agents' design. The engineering experience so far has shown that while proactive agents are harder to design, reactive agents are harder to predict. Consequently it seems that reactive agents are more suitable for developing models, especially in the case of complex systems.

While in the past decade the agent paradigm has significantly evolved, the progress made in modelling complex systems has been far from satisfactory. The existing agent development frameworks have been concentrated on systems engineering. Most of them follow ideas from the OOP paradigm. Only in some cases are concepts like emergence and self organization taken into account and even then with very limited success. The fact is that the problems faced by most of the Agent Oriented Software Engineering (AOSE) methodologies are far more fundamental. Thus there is no complete approach for development of controllable emergent MAS. On the other hand, the inability to deal with emergence creates additional engineering problems for AOSE. Since there is no method to predict emergent phenomena, it is impossible to give guarantees that “negative” emergents won't compromise the system at runtime. By this rationale to be able to make some realistic progress on this problem we need a mechanism that would be able to define and predict what complex MAS would look like at any given time and any desired scale. However given the problems discussed in previous sections as well as our current understanding of emergence, this goal seems completely unrealistic as well.

Consequently at the time being we should concentrate our efforts in defining what type of emergence can (possibly) appear in the system and predict certain features on large time scales with coarse details. This should be done towards defining the causes and effects of the emergent phenomena within the system. In order to find answers to these problems we need to analyze, model and experiment as to identify networks, connection (interaction) patterns and roughly define the relation between the macro and micro levels.

4.3 Micro-Macro Link problem analysis

The emergent behaviour is presumed to exist on the border between the agent interaction at the micro level and system properties at the macro level. Therefore understanding the link between the micro and the macro level is an essential issue in understanding emergence. The quest for understanding the micro-macro link is essentially a quest for achieving controllable and predictable macro level emergents based on micro level interactions. While this is probably the most important question in exploring emergence, it is at the same time the most difficult one.

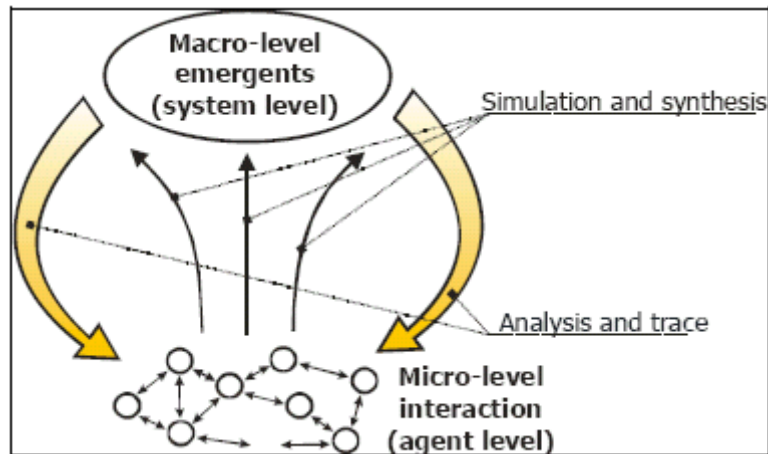


Figure 1: Two way approach in dealing with emergence. Taken from [20]

There has been a variety of attempts to resolve this problem. However most of them are based on theoretical analysis of the problem resulting in general suggestions on how to approach the MML problem. One of these approaches was formulated by Conte and Castelfranchi in [21] where they argued that the micro-macro link requires a two-way approach. Their view encompasses both bottom-up and top-down processes in order to define the micro-macro connections. The bottom-up phase can be done through simulation aimed at attaining collective behaviour from the individual agents. This is the process that defines how the behaviour of individual agents and the communication between them are combined and aggregated. The top-down phase on the other hand is concerned with the feedback that the system at the global level propagates to the agents at the micro level. The top-down analysis is aimed at identification of the relevant micro level properties responsible for the emergents. While putting this approach to practice is harder than it looks it seems like a step in the right direction.

5. Summary and Future work

In recent years the issue of emergence as part of complex systems has gained in importance, partially due to the problems faced by the existing ICT systems as well as the potential for advancing the engineering of adaptive self-organizing systems. Nevertheless the current scientific community has faced many problems dealing with the definition of complex systems, the scope of complexity theory and the types and varieties of emergence.

The phenomenon itself causes additional problems in several ways. It cannot be captured by any model of the system. It is highly unpredictable, and in some cases it is even theoretically unpredictable. In some cases the phenomenon of emergence is associated with extreme jumps in complexity, which is impossible to study. Finally the whole concept is extremely subjective because the emergent behaviour is in the eye of the observer. All these obstacles make any formal approach towards defining the concept of emergence completely unrealistic. Consequently in recent years empirical approaches based on multi-agent simulation are suggested as technique for exploring emergence. Although the efforts to model complex systems (exhibiting emergence) has shown limited success so far, some of the ideas exhibit great potential.

In this context future work includes the development of a framework for the analysis of a specific type of emergent behaviour within complex systems. Unlike most of the attempts, our approach is founded on exploitation of emergent systems already existing in nature (i.e. natural systems) as a basis for building a methodology able to analyze emergent behaviour. The framework itself relies on multi-agent modelling and simulation in order to perform the synthesis of agents' micro level behaviours towards emergent macro level patterns. Then through iterative refinements to the model, the aim is to define the causes and effects of the specific emergent phenomena and thus to delineate the micro-macro relations. While the framework itself can be a useful tool for analysis of complex systems

exhibiting emergence, the knowledge gained with the analysis of the natural system could be used in creating a strategy for engineering emergent behaviours.

References

- 1 Seth B, Dave C. Complexity and emergent behaviour in ICT systems. Technical Report HPL-2004-187, Hewlett Packard; 2004.
- 2 Fortnow L, Homer S. A Short History of Computational Complexity. In D. van D. J Dawson, and A. Kanamori, editors, *The History of Mathematical Logic*. North-Holland, Amsterdam 2002/2003.
- 3 Adami C. What is complexity?. *BioEssays* 2002; 24(12):467–482.
- 4 Badii R, Politi A. Complexity: Hierarchical structures and scaling in physics. Cambridge: Cambridge University Press 1997.
- 5 Holland J H. Hidden Order: How Adaptation Builds Complexity, Addison-Wesley, 1995.
- 6 Herbert S. "The Architecture of Complexity" in *Proceedings of the American Philosophical Society* 1962, 106:467-482.
- 7 Anderson P W, More Is Different, *Science*, New Series 1972, 177(4047):393-396.
- 8 Wolf D T, Holvoet T. Towards a Methodology for Engineering Self-Organising Emergent Systems, In *Self-Organization and Autonomic Informatics (I)*, IOS Press 2005, 135 (5):18-34.
- 9 Chalmers D J. Varieties of Emergence, Preprint, 2002. <http://www.consc.net/papers/granada.html>
- 10 Bedau M. A. Weak emergence Philosophical Perspectives 11:375-399, 1997.
- 11 F. Heylighen, Modelling Emergence, *World Futures: the Journal of General Evolution* 1991, Special Issue on Emergence, 31: 89-104.
- 12 Fromm J. Types and forms of emergence. *Complexity Digest*, 25(3), 2005.
- 13 Wolfram S. Universality and complexity in cellular automata. *Physica D* 198; 1–35.
- 14 Fromm J. The emergence of complexity. Kassel University press 2004.
- 15 Wolf T D, Holvoet T. Emergence versus Self-Organization: Different Concepts but Promising When Combined, *Engineering Self Organising Systems: Methodologies and Applications*, Lecture Notes in Computer Science 2005; 1-15.
- 16 Dumpert K. *The Social Biology of Ants*. Pitman Advanced Publishing, Boston, 1981.
- 17 Seeley T D. *The Wisdom of the Hive: The Social Physiology of Honey Bee Colonies*. Harvard University Press; 1996.
- 18 Wegner P. Why interaction is more powerful than algorithms. *Commun. ACM* 1997; 40(5):80–91.
- 19 Edmonds B, Bryson J. The Insufficiency of Formal Design Methods - the necessity of an experimental approach for the understanding and control of complex MAS. In *Proceedings of the 3rd International Joint Conference on Autonomous Agents & Multi Agent Systems (AAMAS'04)*, New York, ACM Press 2004.
- 20 Fromm J. Ten questions about emergence. *Complexity Digest* 2005; 40(15).
- 21 Conte R., Castelfranchi, C. Simulating multi-agent interdependencies. A two-way approach to the micro-macro link, in Klaus G. Troitzsch et al. (eds.), *Social science microsimulation*, Springer 1995; 394-415.

A Hybrid Formal Approach for Modelling Change in Multi-Agent Systems

Ioanna Stamatopoulou¹, Petros Kefalas², Marian Gheorghe³

¹SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, istamatopoulou@seerc.info

²City College, Dept. of Computer Science, 13 Tsimiski Str, 54624 Thessaloniki, Greece, kefalas@city.academic.gr

³The University of Sheffield, Dept. of Computer Science, Regent Court, 211 Portobello Str, Sheffield S1 4DP, UK, marian@dcs.shef.ac.uk

Biological and biology-inspired systems can be directly mapped to multi-agent systems. By modelling each biological entity as an agent, the overall system behaviour can emerge as the result of the agents' interactions, among them and between them and the environment. One of the most challenging characteristics of such systems that need to be addressed during the modelling phase is the inherent dynamic nature of their structure. New agents may appear in the system, others cease to exist and the communication among them may be constantly changing. Developing such software systems allows us both to observe their behaviour and also inspire and provide new ways for problem solving. Verification and testing techniques, primarily offered by formal methods, are necessary toward increasing the degree of confidence that a system is correct. This paper suggests that two formal methods, Population P Systems and Communicating X-machines, offering complementary advantages, may be combined into one hybrid method, which facilitates the correct specification of multi-agent systems. Population P Systems deal with the dynamic reconfiguration of the system's structure and Communication X-machines facilitates the specification of the agents and their communication. We also demonstrate the practicality of the approach by using an example of a biology-inspired multi-agent system.

Keywords

Formal modelling, Multi-Agent Systems, dynamic structure, X-machines, Population P Systems.

1. Introduction

Throughout the past years, there has been an increasing interest towards biological and biologically inspired systems from various different perspectives within the wider field of computer science. From the smallest living elements, the cells, and how they form tissues in organisms to entire ecosystems and how they evolve, computer scientists are in constant investigation on ways of specifying such systems with the intent of creating software systems that model the behaviour of their biological counterparts. Examples of other biological systems of interest include insect colonies (of ants, termites, bees etc.), flocks of birds, tumours growth—the list is endless.

The motivation behind the development of such software systems varies. To start with, there is the apparent need of biologists to simulate and observe their behaviour, acquire a better understanding of the operations taking place within and the dependencies that influence the ways in which they evolve, or even predict future development. Additionally, however, the understanding of how nature deals with various problematic situations has inspired problem solving techniques that are applicable to a wide range of situations that had been troubling computer scientists. Swarm Intelligence and Ant Colony Optimisation techniques [1], for example, have been successfully applied to robotics [2] and DNA computing [3]. Other unconventional computational models [4] can solve NP-complete problems in non-combinatorial time.

These biology or biologically inspired systems can be directly mapped to multi-agent systems (MAS). Such a claim is not an arbitrary assumption: by considering each biological entity as an agent, with its own behavioural rules, knowledge, decision making mechanisms and means of communication with the other entities and with the environment, the overall system's behaviour is merely the result of the agents' individual actions, the interactions among them and between them and the environment. This also points to the issue of self-organisation and how collective behavioural patterns emerge as a consequence of individuals' local interactions in the lack of knowledge of the entire environment or global control.

There is a number of agent engineering paradigms widely used in industry and academia and naturally each of them has different advantages to offer. Unsurprisingly, the more complex a MAS is, the more difficult the modelling process turns out to be and, in consequence, the less easy it is to ensure correctness at the modelling level that will increase the confidence in the implementation. Correctness implies that all desired properties are verified at the end of the modelling phase and that an appropriate testing technique is applied to prove that the implementation has been built in accordance to the verified model. Not all agent-engineering paradigms provide such means and it is accepted that the most reliable means for ensuring correctness lie within the field of formal methods that can, by nature, provide the necessary verification and testing techniques [5].

Another key aspect that has to be dealt with at the modelling level is the dynamic nature of the configuration of such MAS. Configuration implies the number of the agents, and either their physical placement in space or more generally the structure that is dictated by the communication channels among them. Most modelling methodologies assume a fixed, static structure that is not realistic since in a dynamic MAS communication between two agents may need to be established or ceased at any point and also new agents may appear in the system while existing ones may be removed.

For the purposes of exploring this aspect of MAS we consider the following scenario: a number of identical agents A are located in a plane and move freely (randomly) in space. When two identical agents A collide (Fig. 1.a), a new type of agent L is generated (Fig. 1.b). When any agent A comes close (within a threshold distance) to any agent L, agent A follows the movement of L from then on, as a satellite, staying at the threshold distance (Fig. 1.b). Agents L can have up to a certain number of satellite agents—if this number is reached then a complete assembly is formed. A complete assembly has the ability to immobilise or destroy any agent A, which enters the virtual cycle of the assembly (Fig. 1.c). This kind of scenario resembles a number of situations, which appear in abundance in chemistry, biology, swarms, robotics, artificial life systems, self-assembly etc.

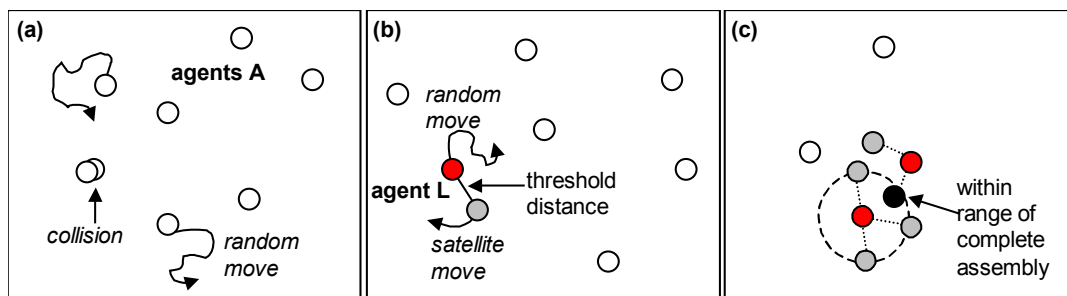


Figure 1 Three system instances showing the agents' behaviour.

In this paper we suggest that a combination of two formal paradigms, namely Population P Systems and Communicating X-machines, can be very promising for the purposes of developing dynamic models of MAS of the nature of many biology and biology-inspired systems. The next section provides a brief background on formal methods and its most popular representatives. Section 3 presents the theory of Population P Systems while Section 4 acquaints the reader with Communicating X-machines and Section 5 presents key issues regarding the reconfiguration of Communicating X-machine systems. Section 6 introduces a hybrid approach for the modelling of a MAS and an example is used in order to portray the potential of the proposed method. Finally, Section 7 discusses issues arising from our attempt and concludes the paper.

2. Formal Methods For Multi-Agent Systems

In an attempt to formally model each individual agent as well as the dynamic behaviour of the overall system, we need a formal method that is capable of rigorously describing all the essential aspects, i.e. knowledge, behaviour, communication and dynamics. It is also important that the level of abstraction imposed by a formal method is appropriate enough to lead towards the implementation of a system. The most widely used formal methods are accompanied by toolkits, which make their adoption wider by researchers and industry. A plethora of formal methods is available for use. Some of them have the means to efficiently define the data structures of a system and the operations employed to modify the values in these structures (Z, VDM). Others are better in describing the control over a system's states (FSM, Petri Nets) and yet others put emphasis on the concurrency and communication of processes (CCS, CSP). Finally, new computation approaches as well as programming paradigms inspired by biological processes in living cells introduce concurrency as well as neatly tackle the dynamic structure of multi-component systems (P Systems, Brane Calculus, Gamma, Cham, MGS) [6], [7], [8].

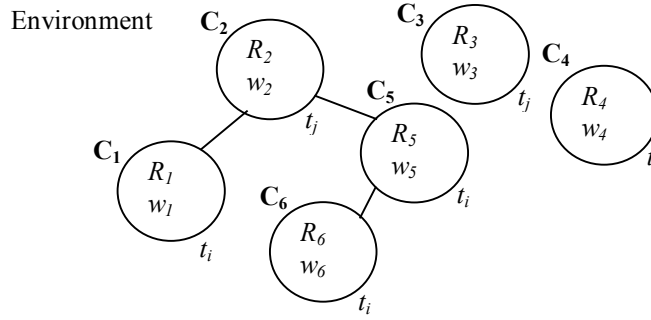
In agent-oriented software engineering, there have been several attempts to use formal methods, each one focusing on different aspects of agent systems development, in order to move to the implementation through refinement of the specification and to be able to develop proof theories for the architecture [9], capture the dynamics of an agent system [10], focus on the specification of the dynamics of the reasoning and acting behaviour of multi-agent systems [11], etc. Other attempts were made in order to verify properties of agent models, based on model checking, or to focus on program generation of reactive systems through a formal transformation process [12], [13]. Wider approaches formally specify multi-agent systems and then directly execute the specification while verifying important temporal properties [14]. Finally, less formal approaches, which accommodate the distinctive requirements of agents, have been proposed [15]. An interesting comparison of various formal methods for the verification of emergent behaviours in swarm-based systems is reported in [16].

The primary aspect of MAS that motivates our work is their dynamic nature in terms of the number of participating agents and the communication channels among them. Population P Systems provide a straightforward way for dealing with the change of a system's structure, however, the rules specifying the behaviour of the individual cells (agents) in a Population P System are more commonly of the simple form of rewrite rules which are not sufficient for describing the behaviour of the respective agent a cell may represent.

On the contrary, X-machines, a state-based formal method introduced by Eilenberg [17], are considered suitable as a specification language of a system's components [5]. Stream X-machines, in particular, were found to be well suited for the modelling of reactive systems. Since then valuable findings using the X-machines as a formal notation for specification, communication, verification and testing purposes have been reported (see [18], [19], [5]).

3. Population P Systems With Active Cells

Population P Systems, a variant of P Systems [6], which were inspired by the processes taking place inside living cells, is a new computational model [20] whereby cells form an arbitrary undirected graph. Each node in the graph represents a membrane, which gets assigned a multi-set of objects. In the absence of a hierarchical membrane structure a Population P System (PPS) is a collection of different types of cells evolving according to specific rules and able of exchanging biological/chemical substances with their neighbouring cells as these are specified by the edges of the graph (Fig. 2). Conversely, other sets of rules are responsible for the reconfiguration of the graph by means of creating or deleting edges between cells and allowing new cells to appear in the system or existing ones to be removed. In addition, an operation of cell differentiation is considered that allows the type of the cells to be changed, thus varying in this way the sets of rules that can be used inside the cells.



- V is a finite alphabet of symbols called objects;
- K is a finite alphabet of symbols, which define different types of cells;
- $\gamma = (\{1, 2, \dots, n\}, A)$, with $A \subseteq \{\{i, j\} \mid 1 \leq i \neq j \leq n\}$, is a finite undirected graph;
- α is a finite set of bond-making rules of the form (t, x_1, x_2, p) , with $x_1, x_2 \in V^*$, and $t, p \in K$;
- $w_E \in V^*$ is a finite multi-set of objects initially assigned to the environment;
- $C_i^p = (w_i, t_i)$, for each $1 \leq i \leq n$, with $w_i \in V^*$ being a finite multi-set of objects, and $t_i \in K$ the type of cell i ;
- R is a finite set of rules dealing with object communication, object transformation, cell differentiation, cell division and cell death.

A short description of the types of rules follows [21]:

Communication rules $(a; b, in)_t$: An object a inside a cell of type t is consumed; an object b is obtained by a neighbouring cell.

$(a; b, enter)_t$: An object a inside a cell of type t is consumed and an object b is obtained by the environment.

$(b, exit)_t$: An object b inside a cell of type t is expelled out into the environment

Transformation rules $(a \rightarrow b)_t$: An object a is replaced by an object b within a cell of type t .

Cell differentiation rules $(a)_t \rightarrow (b)_p$: An object a is replaced by an object b and the type of the cell changes to p .

Cell division rules $(a)_t \rightarrow (b)_t (c)_t$: A cell of type t containing an object a is divided into two cells of the same type. One of the new cells has a replaced by b while the other by c . All other objects of the originating cell appear unaltered in both new cells.

Cell death rules $(a)_t \rightarrow \dagger$: An object a inside a cell of type t causes the removal of the cell from the system.

Bond-making rules (t, x_1, x_2, p) : If an object x_1 exists in a cell of type t and an object x_2 in a cell of type p they are consumed and a bond (edge) is created between the two cells.

These networks of communicating membranes can be interpreted as an abstract model of bio-entities aggregated together in more complex bio-units. In this respect the model may also address the cases of various colonies of more complex organisms like ants, bees etc. and can be considered as a method for modelling MAS. Assuming that the objects inside each cell represent the knowledge of the respective agent, the transformation and communication rules take up the task of updating and communicating knowledge while the cell rules dictate when new agents should appear in the system or existing ones

should be removed, and the bond-making rules are responsible for establishing communication channels between two agents.

An initial attempt to use solely Population P Systems for the modelling of a MAS [22] has indicated that although a PPS can deal with the reconfiguration of the overall system structure in a straightforward way, it is not as intuitive in specifying each of the participating agents in terms of their knowledge, actions and control over their internal states. The next section presents X-machines and how they may be used for the efficient modelling of the individual agents.

4. Communicating X-Machines

X-machines (XM) possess characteristics that make them useful for specifying software systems. An XM model consists of a number of states, just as a Finite State Machine does, but in contrast to that, an XM model has a memory, which accommodates mathematically defined data structures, pretty much as Z does. The transitions between states are labelled by functions, which are not applied only to inputs but also to memory values and produce outputs and new memory values.

Definition 2 The 8-tuple $XM = (\Sigma, \Gamma, Q, M, \Phi, F, q_0, m_0)$ defines a *stream X-machine* [5] where:

- Σ and Γ are the input and output alphabets respectively;
- Q is the finite set of states;
- M is the (possibly) infinite set called memory;
- Φ is a set of partial functions φ that map an input and a memory state to an output and a possibly different memory state, $\varphi: \Sigma \times M \rightarrow \Gamma \times M$;
- F is the next state partial function, $F: Q \times \Phi \rightarrow Q$, which given a state and a function from the type Φ determines the next state. F is often described as a state transition diagram;
- q_0 and m_0 are the initial state and initial memory respectively.

The machine consumes a stream of inputs and produces a stream of outputs, through a number of computation steps (a computation step being the application of one function) and computation ceases when there are no more inputs to be consumed.

Definition 3 A computation step is defined as a transition of an XM. A transition is the change $(m, q, s, g) \vdash (m', q', s', g')$ where $m, m' \in M$, $q, q' \in Q$, the input stream $s = \sigma s'$, $\sigma \in \Sigma$, the output stream $g' = g\gamma$, $\gamma \in \Gamma$ and $\exists \varphi \in \Phi: q' \in F(q, \varphi)$, $\varphi(\sigma, m) = (\gamma, m')$.

An XM with no initial state and memory is called an X-machine type. Types can be used to create instances of XM that can all be part of a larger system so long as there are means for communication among them. XM instances that are able to communicate are called Communicating X-machines components (CXM). Though a number of approaches have been proposed for both asynchronous and synchronous communication between the machines [18], in principle, communication is established by directing the output of a function of a machine to the input stream of another machine.

This allows the modelling of more complex systems in a bottom-up approach: since a *Communicating X-machine system*, Z , consists of several CXM components that are able to exchange messages, modelling a MAS involves initially specifying the types of participating agents as XMs, creating the required number of instances (CXM components C_i) and later defining the communication channels among them as a relation $CR \subseteq C \times C$, where $C = \{C_1 \dots C_n\}$ is the set of n agent instances present in the system. A tuple (C_i, C_k) denotes that the CXM component C_i can output a message to a corresponding input stream of CXM component C_k for any $i \neq k$. In consequence, the overall system structure is defined as the graph whose nodes are the CXM components and edges the communication channels among them.

5. Redefining The Structure Of A System Of CXMs

In order to be able to dynamically redefine the structure of a CXM system model, we are in need of operators that will have an effect on the graph representation of the system structure by reconfiguring the communication channels.

The Attachment operator ATT is responsible for establishing communication between two existing CXM components. It takes as arguments two CXM components C_i , C_k and the current communicating system (Z) to which they belong and updates the system (Z') so that C_i and C_k are able to communicate. The Detachment operator DET removes communication channels between two existing CXM components in Z . The Generation operator GEN creates and introduces a new CXM component (of a particular type) into the updated communicating system Z' . The Destruction operator DES is used for the removal of an existing CXM component from the communicating system Z along with all the channels that allow its communication with other components. The interested reader is directed to [23] for the formal definitions.

The rules that drive the evolution of the system structure are generally of the form *condition* \Rightarrow *action* whereby, if the condition allows, an appropriate action, which includes one or more reconfiguration operations, is being performed.

The condition should allow us to express properties regarding the computation state of one or more of the components. For a clear view of the computation of an XM, three pieces of information are required: its current state, its current memory, and the last function that has been applied.

Definition 4 The *computation state* S_i of a CXM C_i is defined as the 3-tuple $S_i = (q_i, m_i, \phi_i)$ where q_i is the state in which C_i is in, m_i is the memory value of C_i and ϕ_i is the last function that has been applied in C_i .

6. A Hybrid Approach

The idea behind combining Population P Systems and Communicating X-machines is that the first formalism will be responsible for the reconfiguration of the modelled system's structure while the latter for the specification of the individual components. Both systems have the same physical structure at any time of the computation, i.e. their graph representation remains the same and each of the PPS cells correspond in a one-to-one manner to the XM components. In order to achieve that, the objects of each of the cells of the PPS need to capture information regarding the computation state of the corresponding CXM. Figure 3 depicts the notion of a PPS working on top but also in parallel and in complete correspondence with a CXM system.

Regarding the scenario that was presented in the introduction, the two types of agents, A and L can be modelled as X-machines, whose state transition diagrams are shown in Fig. 4. The memory of agent A holds its current position, the identifier of the agent L that is being followed (or *noL* if none is followed) and the threshold distance, under which a bond is made with an agent L so $M_A = (Z \times Z) \times (\mathcal{L} \cup \{noL\}) \times R$, where \mathcal{L} is the set of all possible identifiers of agents of type L . The input set is $\Sigma_A = (\mathcal{L} \cup \mathcal{A} \cup \{space\}) \times (Z \times Z)$, where \mathcal{A} is the set of all possible identifiers of agents of type A . The output set Γ_A is a set of messages.

Accordingly, for the agent L , $Q_L = \{moving_freely\}$, $\Sigma_L = (\mathcal{A} \cup \{space\}) \times (Z \times Z)$, and Γ_L is a set of messages. The memory $M_L = (Z \times Z) \times P(\mathcal{A}) \times N$, where the second memory position holds the set of satellite agents of type A and the third position holds the number of agents A that are needed for a complete assembly. Indicatively, some of the functions in the two Φ sets are:

$$\begin{aligned} move((space, (x, y)), ((cx, cy), noL, d)) = \\ (movingFreely, ((x, y), noL, d)), \text{ if } neighbours((x, y), (cx, cy)) \\ follow((myL, (x, y)), ((cx, cy), myL, d)) = \\ (followL, ((cx', cy'), myL, d)), \\ \text{ where } (cx', cy') = calculate_coord(d, (x, y), (cx, cy)) \end{aligned}$$

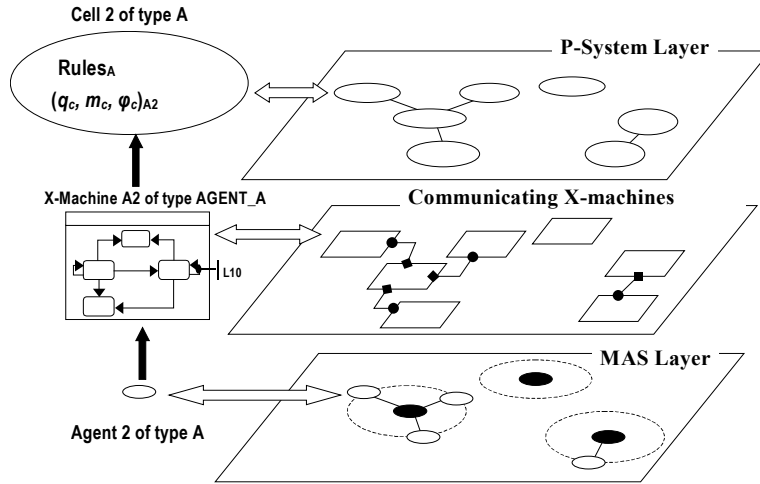


Figure 3 The parallel computation of the Communicating X-machine system and the Population P System that model the behaviour of the MAS.

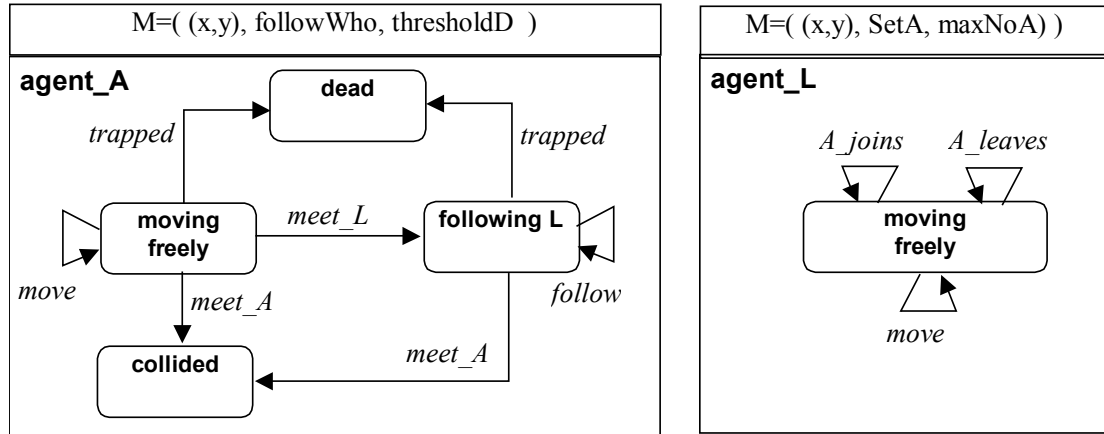


Figure 4 The two X-machine instances used in the initial configuration of the example. The *move* function of L_{10} sends its output to be read by the *follow* function of A_1 .

$move((space, (x, y)), ((cx, cy), setA, maxA)) =$
 $(moving_freely, ((x, y), setA, maxA)),$
 if $neighbours((x, y), (cx, cy))$
 $A_joins((agentA, (x, y), ((cx, cy), setA, maxA)) =$
 $(agentAJoins, ((cx, cy), setA \cup agentA, maxA)),$
 if $|setA| < maxA \wedge agentA \notin setA$

Let us assume an there is one agent instance of type L , L_{10} , and one agent instance of type A , A_1 , being a satellite of L_{10} . While moving, L_{10} sends its new coordinates to A_1 , which uses them as input to follow L_{10} . The corresponding function of L_{10} is extended with a transformation operation so that its output is changed into the format that is understandable by the follow function of the A agent.

This provides a static configuration of two agents that may be considered as an initial state of the system. In order for the model to be able to dynamically restructure itself we also define a PPS comprised of two cells:

$P_{satellites} = (V, K, \gamma, \alpha, \emptyset, C_{A1}^p, C_{L10}^p, R)$ where the alphabet V will be the set of possible computation states of all the CXMs, the set of cell types is $K = \{L, A\}$ and the graph $\gamma = \{(C_{L10}^p, C_{A1}^p)\}$.

The set of rules that handle the reconfiguration of this hybrid system are defined to be of the form $PPS_rule \Rightarrow CXM_action$, meaning that the condition for a reconfiguration operator to be applied at the CXM system level is the existence and application of a corresponding PPS rule.

Baring the above in mind, the set α of bond-making rules will contain one rule as the only communication established is between an L agent and an A agent that becomes L 's satellite:

$$(A, (moving_freely, ((x1, y1), noL, d), \varphi_{Ai}) ; (moving_freely, ((x2, y2), setA, max), \varphi_{Lj}), L) \\ \Rightarrow Z' = \mathbf{ATT}(Ai, Lj, Z) \\ \text{if } d \leq distance((x1, y1), (x2, y2)) \wedge |setA| < m$$

There is no need for communication rules as communication is being handled by the CXM system itself.

The transformation rule will be of the same form for all types of cells. Since the only object in each cell is the 3-tuple that denotes the computation state of the corresponding XM, then this tuple is updated according to the computation step performed by the XM:

$$((q_{Li}, m_{Li}, \varphi_{Li}))_L \xrightarrow{XM_{Li}} ((q'_{Li}, m'_{Li}, \varphi'_{Li}))_L$$

where q'_{Li}, m'_{Li} are derived from the computation step of XM_{Li} when in state q_{Li} with memory m_{Li} and it consumes an input from its input stream.

A cell differentiation rule will be responsible for defining that an agent A that has collided with another agent A will be transformed into a leader. A cell death rule ensures that the other collided agent A disappears from the system.

$$((collided, ((x, y), _, _), \varphi_{Ai}))_A \rightarrow ((moving_freely, ((x, y), \emptyset, max), \epsilon))_L \\ \Rightarrow Z' = \mathbf{GEN}(L, \mathbf{DES}(A_i, Z)) \\ ((dead, M_{Ai}, \varphi_{Ai}))_A \rightarrow \dagger \\ \Rightarrow Z' = \mathbf{DES}(A_i, Z)$$

7. Conclusions

This work has been motivated by the need for a formal framework for capturing the dynamics of the structure of biology or biology-inspired MAS during their modelling phase. We consider a Population P System being in charge of the reconfiguration of such a system, whereby each cell of the PPS corresponds to an agent of the MAS. Since PPSs are not as intuitive when it comes to the modelling of the individual agents, Communicating X-machines are used for this purpose. They are intuitive and can easily capture both the data and the control over the internal states of an agent.

XMs provide a solid mathematical framework and can therefore also support the formal verification of desired properties that should hold for each agent through the use of a model checking technique [19]. Thus, in relation to our example, properties such as "an agent L can never have more than the maximum number of allowed satellites" can be verified before using the agent's model in a CXM system. Furthermore, XMs offer a testing strategy that, under certain conditions, grants to find all faults in an implementation of a model [5].

Our work suggests a hybrid approach whereby a PPS and CXM system work in parallel, the first taking care of the dynamic aspects of a MAS and the second providing the means for the specification of the individual agents and the communication among them. Future work primarily includes concluding the formal definition of this hybrid model. Additionally, efforts will be directed towards implementing those features on top of existing XM and PPS animators and integrating the latter in order to be able to validate the practicality of the approach. Finally, we are experimenting with the automatic transformation of XM models written using a particular notation into NetLogo [24] so that we have a straightforward visualisation of the models.

References

- 1 Dorigo M, Maniezzo V, Colomi A. The Ant System: optimisation by a colony of cooperating agents. IEEE Transactions on Systems, Man and Cybernetics. 1996; 26(1):1–13.

- 2 Dorigo M, Trianni V, Sahin E, Gross R, Labella TH, Baldassarre G, et al. Evolving Self-Organizing Behavior. *Autonomous Robots*. 2004; 17(2-3):223–245.
- 3 Păun G, Rozenberg G, Salomaa A. *DNA Computing: New Computing Paradigms*. Springer-Verlag; 1998.
- 4 Gheorghe M, editor. *Molecular Computational Models: Unconventional Approaches*. Idea Publishing Inc.; 2005.
- 5 Holcombe M, Ipaté F. *Correct Systems: Building a Business Process Solution*. London: Springer-Verlag; 1998.
- 6 Păun G. Computing with membranes. *Journal of Computer and System Sciences*. 2000; 61(1):108–143.
- 7 Banatre JP, Le Metayer D. The gamma model and its discipline of programming. *Science of Computer Programming*. 1990; 15:55–77.
- 8 Berry G, Boudol G. The chemical abstract machine. *Journal of Theoretical Computer Science*. 1992;96(1):217–248.
- 9 d’Inverno M, Kinny D, Luck M, Wooldridge M. A formal specification of dMARS. In: Singh MP, Rao A, Wooldridge MJ, editors. *Intelligent Agents IV*. vol. 1365 of *Lecture Notes in AI*. Springer-Verlag; 1998; p. 155–176.
- 10 Rosenschein SR, Kaebling LP. A situated view of representation and control. *Artificial Intelligence*. 1995; 73:149–173.
- 11 Brazier F, Dunin-Keplicz B, Jennings N, Treur J. Formal specification of multi-agent systems: a real-world case. In: *Proceedings of International Conference on Multi-Agent Systems (ICMAS’95)*. MIT Press; 1995; p. 25–32.
- 12 Benerecetti M, Giunchiglia F, Serafini L. A model-checking algorithm for multi-agent systems. In: Muller JP, Singh MP, Rao AS, editors. *Intelligent Agents V*. *Lecture Notes in Artificial Intelligence*. Springer-Verlag; 1999; p. 163–176.
- 13 Attoui A, Hasbani A. Reactive systems developing by formal specification transformations. In: *Proceedings of the 8th International Workshop on Database and Expert Systems Applications (DEXA 97)*; 1997; p. 339 – 344.
- 14 Fisher M, Wooldridge M. On the formal specification and verification of multi-agent systems. *International Journal of Cooperating Information Systems*. 1997; 6(1):37–65.
- 15 Odell J, Parunak HVD, Bauer B. Extending UML for agents. In: *Proceedings of the Agent-Oriented Information Systems Workshop at the 17th National Conf. on Artificial Intelligence*; 2000; p. 3–17.
- 16 Rouf C, Vanderbilt A, Truszkowski W, Rash J, Hinchey M. Verification of NASA emergent systems. In: *Proceedings of the 9th IEEE International Conference on Engineering Complex Computer Systems (ICECCS’04)*; 2004. p. 231–238.
- 17 Eilenberg S. *Automata, Languages and Machines*. Academic Press; 1974.
- 18 Kefalas P, Eleftherakis G, Kehris E. Communicating X-machines: a practical approach for formal and modular specification of large systems. *Journal of Information and Software Technology*. 2003;45(5):269–280.
- 19 Eleftherakis G. *Formal Verification of X-machine Models: Towards Formal Development of Computer-based Systems*. PhD Dissertation, Department of Computer Science, University of Sheffield; 2003.
- 20 Bernardini F, Gheorghe M. Population P Systems. *Journal of Universal Computer Science*. 2004;10(5):509–539.
- 21 Kefalas P, Stamatopoulou I, Gheorghe M. A formal modelling framework for developing multi-agent systems with dynamic structure and behaviour. In: Pechoucek M, Petta P, Varga LZ, editors. *Proceedings of the 4th International Central and Eastern European Conference on Multi-Agent Systems, CEEMAS 2005*. No. 3690 in *Lecture Notes in Artificial Intelligence*. Springer-Verlag; 2005; p. 122–131.
- 22 Stamatopoulou I, Kefalas P, Gheorghe M. Modelling the dynamic structure of biological state-based systems. In: *6th International Workshop on Information Processing in Cells and Tissues*; 2005.

- 23 Kefalas P, Eleftherakis G, Holcombe M, Stamatopoulou I. Formal Modelling of the Dynamic Behaviour of Biology-Inspired Agent-Based Systems. In: Gheorghe M, editor. *Molecular Computational Models: Unconventional Approaches*. Idea Publishing Inc.; 2005; p. 243–276.
- 24 Wilensky U. NetLogo. Center for Connected Learning and Computer-based Modelling. Northwestern University, Evanston, IL. <http://ccl.northwestern.edu/netlogo>. 1999.

Modern Applications of Machine Learning

George Tzanis, Ioannis Katakis, Ioannis Partalas, Ioannis Vlahavas

*Department of Informatics, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece,
{gtzanis, katak, partalas, vlahavas}@csd.auth.gr*

Machine learning is one of the older areas of artificial intelligence and concerns the study of computational methods for the discovery of new knowledge and for the management of existing knowledge. Machine learning methods have been applied to various application domains. However, in the few last years due to various technological advances and research efforts (e.g. completion of the Human Genome Project, evolution of the Web), new data have been available and consequently new domains where machine learning can be applied have been arisen. Some of these modern applications are learning from biological sequences, learning from email data, and learning in complex environments such as Web. In this paper we present the above three application domains as well as some recent efforts, where machine learning techniques are applied in order to analyze the data provided by these domains.

Keywords

Bioinformatics, Learning from Email, Machine Learning, Reinforcement Learning.

1. Introduction

A cognitive system tries to understand the concepts of its environment by using a simplified interpretation of this environment called *model*. The procedure of constructing such a model is called *inductive learning*. Moreover, a cognitive system is able to organize its experience by constructing new structures called *patterns*. The construction of models and patterns by a cognitive system using a dataset is called *machine learning*. Machine learning tasks can be classified into the following two groups:

- Supervised learning
- Unsupervised learning

A model describes the whole set of data and is also characterized as *predictive model* since it can be used to predict the output of a function (*target function*) for a given value in the function's domain. Moreover, a model provides some qualitative information about the data. In contrast, a pattern describes only a portion of the data and is characterized as *informative pattern*.

1.1 Supervised Learning

This kind of learning is also known as *learning from examples*. In supervised learning the cognitive system has to learn a concept or a function that is actually the description of a model. In particular the system is provided with a set of examples. The output of the target function for each of these examples is also available. The system has to discover the description of the model based on the output of the function. For evaluation purposes a model is built using a subset of the data (*training set*), while the remaining data are used to evaluate the constructed model (*test set*).

Two learning tasks are recognized in supervised machine learning, namely *classification* and *regression*. Classification concerns the construction of prediction models for functions with discrete range, while regression concerns the construction of prediction models for functions with continuous range. The most common supervised machine learning methods are the following:

- *Concept Learning.* The cognitive system is provided with examples that belong (positive examples) or do not belong (negative examples) in a concept (class). Then, the system is called to produce a generalized description of the concept in order to decide for future cases based on this description.
- *Classification or Decision Tree Induction.* Classification or decision tree induction methods are very popular and are used for the approximation of discrete target functions. These methods construct tree structures that represent graphically the training data. The main advantage of decision trees is that they are easily interpreted. Decision trees can also be represented as “if-then” rules.
- *Rule Learning.* Rule learning includes the induction of “if-then” rules, called *classification rules*. Classification rules are used for the approximation of discrete target functions.
- *Instance Based Learning.* In this kind of learning the data are stored in their raw format. When the system is called to decide for a new case it examines the relationship between the new case and each of the stored examples. This kind of learning is also known as *lazy learning*, since the learning process is deferred until a new case appears.
- *Bayesian Learning.* This kind of learning is based on Bayes theorem and includes methods that utilize probabilities. Existing knowledge can be incorporated in the form of initial probabilities.
- *Linear Regression.* Linear regression is a method for describing a target function with a linear combination of a number of other variables. The target function’s range must be a continuous interval.
- *Neural Networks.* Neural networks can be used both for classification and regression. Their function is based on biological patterns and various procedures that simulate the human brain’s activity are used.
- *Support Vector Machines (SVMs).* SVMs are based on statistical learning theory and neural networks. SVMs are very popular in classification and regression tasks and usually present very good prediction accuracy.

1.2 Unsupervised Learning

This kind of learning is also known as *learning from observation*. In unsupervised learning the system has to discover any patterns (i.e. associations or clusters) based only on the common properties of the example without knowing how many or even if there are any patterns. The main unsupervised machine learning methods are the following:

- *Association Rule Mining.* Association rule mining emerged in 1993 [1] and has many contributions from the research area of databases. It was introduced as a market basket analysis method. Association rules are implications of the form $A \Rightarrow B$. The interpretation of the above rule is that when item A appears in a basket, then item B will also appear in the same basket.
- *Sequential Pattern Mining.* Sequential pattern mining concerns the learning from ordered data. The order is usually temporal. It has also many contributions from the research area of databases and has been proposed [2] as an extension of association rules mining.
- *Clustering.* Clustering is the procedure of discovering clusters of examples, so that examples that belong to the same cluster are as similar as possible, while examples belonging in separate clusters are as dissimilar as possible.

1.3 Applications

Machine learning has been extensively applied in various application domains. Some of the most popular applications include medical diagnosis, credit risk analysis, customer profiling, market

segmentation, targeted marketing, retail management and fraud detection. The last years due to various technological advances and research efforts like the completion of the Human Genome Project (http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml) and the evolution of the Web, new data have been available and consequently new domains where machine learning can be applied have been arisen. Some of these modern applications are learning from biological sequences, learning from text, and learning in complex environments such as Web. The above three modern applications of machine learning are presented below. In particular, in section 2 the application of machine learning on biological sequences is presented, section 3 deals with learning from text and section 4 concerns focused crawling using reinforcement learning. Finally, section 5 concludes the paper.

2. Learning from Biological Sequences

The rapid progress of computer science in the last decades has been closely followed by a similar progress in molecular biology. Undoubtedly, the use of computational tools has given a boost in the collection and analysis of biological data, creating one of the hottest areas of research, namely *bioinformatics*. Biological sequences represent a large portion of the biological data that require the use of computational tools in order to be analyzed. The large size of the sequences and the numerous possible features are the main reasons behind the urgent need methods that allow for the efficient analysis of such data and the delivery of accurate and reliable knowledge to the domain expert. The field of machine learning provides the biologists with a big set of tools for the analysis of these data.

Many machine learning techniques have been proposed to deal with the identification of specific biological sequence segments. The most common include neural networks, Bayesian classifiers, decision trees, and Support Vector Machines [3, 4, 5]. Sequence recognition algorithms exhibit performance tradeoffs between increasing sensitivity (ability to detect true positives) and decreasing selectivity (ability to exclude false positives) [6]. However, as Li et al. [7] state, traditional machine learning techniques cannot be directly applied to this type of recognition problems. Thus, there is the need to adapt the existing techniques to this kind of problems. Attempts to overcome this problem have been made using feature generation and feature selection [7, 8]. Another machine learning application is the use of clustering algorithms to group structurally related biological sequences.

2.1 Prediction of Translation Initiation Sites

The main structural and functional molecules of an organism's cell are *proteins*. The information concerning the synthesis of each protein is encoded by the genetic material of the organism. The genetic material of almost every living organism is *DNA*. Another molecule that plays an important role in protein synthesis is *RNA*. DNA and RNA belong to a family of molecules called *nucleic acids*. Both proteins and nucleic acids are sequences of smaller molecules, *amino acids* and *nucleotides* respectively. A sequence can be represented as a string of different symbols. There are twenty amino acids and five nucleotides. Every nucleotide is characterized by one of the following letters: A, C, G, T, U. DNA may contain a combination of A, C, G, and T. In RNA U appears instead of T. Proteins are synthesized by the following process. DNA is used as template for the synthesis of RNA (*transcription*). Then RNA is used as template for the synthesis of a protein molecule (*translation*).

Translation takes place by an organelle called *ribosome*. The mRNA sequence is scanned by the ribosome, which reads triplets, or *codons*, of nucleotides and “translates” them into amino acids. Thus, a protein consisting of n amino acids is encoded by a sequence of $3n$ nucleotides. Since there are 64 different triplets formed from an alphabet of four nucleotides and the total number of amino acids is 20, it is obvious that some amino acids are encoded by more than one codon. Moreover, the triplet AUG, that encodes amino acid methionine is also used as a translation initiation codon. Finally, there are three stop codons for the termination of translation (UAG, UAA and UGA).

Translation, usually, initiates at the AUG codon nearest to the start of the RNA sequence. However this is not always the case, since there are some escape mechanisms that allow the initiation of translation at following AUG codons. Due to these mechanisms the recognition of the Translation

Initiation Site (TIS) on a given sequence becomes more difficult. After the initiation of translation, the ribosome moves along the RNA molecule, towards the end of the sequence and reads the next codon. This process is repeated until the ribosome reaches a stop codon. For each codon read, the proper amino acid is brought to the protein synthesis site and is joined to the protein chain, which by this way is elongated.

The recognition of the TIS is essential for better understanding of the process of translation. It has been recognized as one of the most critical problems in molecular biology that requires the generation of classification models, in order to accurately and reliably distinguish the valid TISs from a set of false ones.

Although many approaches have been proposed to deal with this problem, there is a great potential for the improvement of their accuracy. In [10] we apply machine learning methods to tackle the problem of the prediction of TISs in DNA sequences. We use a large number of features and different algorithms in order to build more accurate models. Some of the features are directly extracted from the raw sequences, concerning the nucleotides present at each position of the sequence, but most of them are generated. Along with the features already discussed in other papers [5, 8], we have generated and proposed the use of some new ones (*up-down_x*, *up_pos_k_x*, *down_pos_k_x* in Table 1). We have shown that a combination of these features improves the accuracy of the prediction models. In [11] we have presented an extension of [10], where a step of grouping the sequences according to criteria based on the sequence length was incorporated. Moreover, instead of nucleotide pattern, amino acid patterns were used as features. Finally, a multiple classifier system was built. For our experiments we used a real world dataset that contains processed DNA sequences collected from vertebrate organisms [12].

Feature	Description
up_ <i>x</i>	Counts the number of occurrences of amino acid <i>x</i> in the upstream region
down_ <i>x</i>	Counts the number of occurrences of amino acid <i>x</i> in the downstream region
up-down_ <i>x</i>	Counts the difference between the number of occurrences of amino acid <i>x</i> in the upstream region and the number of occurrences of amino acid <i>x</i> in the downstream region
up_pos_ <i>k_x</i>	Counts the number of occurrences of nucleotide <i>x</i> in the <i>k</i> th position of the upstream codons ($k \in \{1, 2, 3\}$)
down_pos_ <i>k_x</i>	Counts the number of occurrences of nucleotide <i>x</i> in the <i>k</i> th position of the downstream codons ($k \in \{1, 2, 3\}$)
up_-3_[AG]	A Boolean feature that is true if there is an A or a G nucleotide three positions before the AUG codon, according to Kozak's pattern (GCC[AG]C C aug G) [9]
down_+1_G	A Boolean feature that is true if there is a G nucleotide in the first position after the AUG codon, according to Kozak's pattern (GCC[AG]C C aug G)
up_AUG	A Boolean feature that is true if there is an in-frame upstream AUG codon
down_stop	A Boolean feature that is true if there is an in-frame downstream stop codon

Table 1 The features used in our approach [11].

In the following lines we describe the approach we have followed in [11] in order to construct a multiple classifier system for the prediction of TISs in genomic sequences. Our approach consists of a number of steps. Each of these steps is described in detail in the following lines.

Step 1: All sequences are scanned and every candidate TIS is detected.

Step 2: The candidate TISs found in step 1 are grouped according to the length of the sequence before the AUG codon (upstream) and after the AUG codon (downstream). By this way the initial dataset of candidate TISs is divided into a number of smaller datasets. In our setup we have divided the initial dataset in 4 smaller datasets (This step was absent from our approach in [10])

Step 3: For each of the candidate TISs the value of a number of features is calculated. More details about these features are listed in Table 1.

Step 4: The features are evaluated among the instances of every group according to their impact in the accuracy of classification. In our setup we have used the information gain measure.

Step 5: A number of the top ranked features is selected and a classifier is built for each of the data subsets.

Finally, a new instance, namely a new candidate ATG, is assigned to one of the groups according to the length of its upstream and downstream regions' length and is classified by the corresponding classifier.

Our approach has been tested using various classification algorithms (e.g. C4.5, Naïve Bayes, PART, RIPPER) and presented better classification accuracy than other approaches.

3. Learning from Email Data

Email has met tremendous popularity over the past few years. People are sending and receiving many messages per day, communicating with partners and friends, or exchanging files and information. Unfortunately, the phenomenon of email overload has grown over the past years becoming a personal headache for users and a financial issue for companies. In this section, we will discuss how Machine Learning can contribute to the solution of this problem [13].

3.1 Automatic Answering

Large companies usually maintain email centres (in conjunction with “call centres”) with employees committed to answer incoming messages. Those messages usually come from company clients and partners and many times address the same problems and queries. Automatic email answering is an effort to build email centers or personalized software that will be able to analyse an incoming message and then propose or even send an applicable answer. Efforts towards this direction have been made recently [14, 15].

3.2 Automatic Mail Organization into Folders

The growth of email usage has forced users to find ways to organize archive and manage their emails more efficiently. Many of them are organizing incoming messages into separate folders. Folders can be topic-oriented like “work”, “personal” and “funny”, people-specific like “John” and “Mary” or group-of-people-specific like “colleagues”, “family” and “friends”. Some users are archiving their messages according to importance and thus maintain folders like “urgent”, “for future reference”, “spam” etc. To achieve this, many users create manually some so-called *rules* to classify their email.

What Machine Learning has to offer to this task is the automatic classification of incoming email by observing past and current classifications made by the user (e.g. analyzing already existing folders or taking a current classification as an example). Thus, the user does not need to create the rules by himself. Furthermore, machine learning algorithms are able to classify a message, taking under consideration its content by searching for specific keywords. This is usually achieved by combining statistical and linguistic techniques. It is extremely convenient for the user, since there are some concepts like “messages concerning my work” or “interesting messages” or “messages that I have to answer today” that cannot easily be described with a combination of keywords. Moreover, these concepts may change (e.g. the concept of “interesting message”) from time to time. A Machine Learning algorithm can learn to classify new messages just by silently observing past examples and can follow drift of concepts by accepting user feedback.

A lot of research has been recorded in the field [16, 17] and lots of those ideas have been implemented into useful email tools [18, 19].

3.3 Email and Thread Summarization

There is a certain category of email users that receive hundreds of messages per day. Some of them are newsletters, others are business decision-making messages from colleagues, appointment arrangements etc. It would be extremely useful for them if they could avoid reading all of those messages and instead read only the most important and necessary parts and then decide if the messages demand immediate attention. From a summary, they could also find out if a newsletter for example is interesting for them or not and only then read the full text. Again, data mining techniques are explored in order to build trainable tools for summarization. [20].

3.4 Spam Filtering

The main goal of spam filtering is to identify and sort out unsolicited commercial mails (spam) from a user's email stream. Spam email has begun as small annoyance in the early days of email to become a major industry problem in the last five years. The large amount of spam not only causes bandwidth (and therefore financial) problems, but also takes up valuable time from email users who try to separate and delete many unsolicited messages every day. Moreover, many spam messages include pornographic content inappropriate for children. Many different machine learning classifiers have been tested in the bibliography including Naïve Bayes [21], Support Vector Machines [22], Stacking Classifiers [23] and some of them have proved to be particularly accurate.

3.6 Dynamic Feature Space and Incremental Feature Selection for Email Classification

Email classification problems are of special interest for the Machine Learning and Data Mining community, mainly because they introduce and combine a number of special difficulties. They deal with high dimensional, streaming, unstructured, and, in many occasions, concept drifting data. Another important peculiarity of email (and streaming text in general), not adequately discussed in the relative literature, is the fact that the feature space is initially unavailable. In this section we present a computationally undemanding method that tackles with this problem [24].

Our approach uses two components in conjunction: a) an incremental feature ranking method, and b) an incremental learning algorithm that can consider a subset of the features during prediction. Feature selection methods that are commonly used for text classification are filters that evaluate the predictive power of each feature and select the N best. Such methods evaluate each word based on cumulative statistics concerning the number of times that it appears in each different class of documents. This renders such methods inherently incremental: When a new labeled document arrives, the statistics are updated and the evaluation can be immediately calculated without the need of re-processing past data. These methods can also handle new words by including them in the vocabulary and initializing their statistics. Therefore the first component of our approach can be instantiated using a variety of such methods, including information gain, the χ^2 statistic or mutual information.

The incremental re-evaluation and addition of words will inevitably result into certain words being promoted to / demoted from the top N words. This raises a problem that requires the second component of the proposed approach: a learning algorithm that is able to classify a new instance taking into account different features over time. This problem has not been considered before to the best of our knowledge. We call learning algorithms that can deal with it *feature-based*, because learning is based on the new subset of features, in the same way that in instance based algorithms, learning is based on the new instance. An inherently feature based algorithms is Naïve Bayes (NB) where, each feature makes an independent contribution towards the prediction of a class. Therefore, it can be easily expanded in order to instantiate the second component of our approach. Specifically, when NB is used for the classification of a new instance, it should also be provided with an additional parameter denoting the subset of the selected features it will then only consider the calculated probabilities of this subset. Figure 1 presents algorithm Update for the incremental update of our approach.

```

input  : Document, DocClass, Classes, Vocabulary
output: Classifier, Vocabulary, WordStats, Evaluation
begin
  foreach Word ∈ Document do
    if Word ∉ Vocabulary then
      ADDWORD(Word, Vocabulary)
      foreach Class ∈ Classes do
        WordStats [Word][Class][1] ← 0
        WordStats [Word][Class][0] ← 0
      end
    end
  end
  foreach Word ∈ Vocabulary do
    if Word ∈ Document then
      WordStats [Word][DocClass][1] ← WordStats [Word][DocClass][1] + 1
    else
      WordStats [Word][DocClass][0] ← WordStats [Word][DocClass][0] + 1
    end
  end
  foreach Word ∈ Vocabulary do
    Evaluation ← EVALUATEFEATURE(Word, WordStats)
  end
  Classifier ← UPDATECLASSIFIER(Document, DocClass)
end

```

Figure 1 Algorithm for the incremental update.

4. Focused Crawling Using Reinforcement Learning

The World Wide Web can be considered the greatest library of all kinds of information that exists. In contrast with other kind of libraries, web lacks indexing structure, which allows the users to access the desired information. In order to deal with this problem search engines have been developed that try to organize and index web pages into extensive catalogues. Search engines try to collect as many pages as possible in order to cover the majority of the thematic topics. This process requires the use of programs called crawlers.

The architecture of these programs is simple. Starting from a set of seed pages, they crawl the web following the hyperlinks included there. Some initial approaches used graph based techniques like depth first search and breadth first search. Due to the explosive growth of the web the crawling that utilizes the above implementations became time and resource consuming process. Moreover, user groups that are interest in specific topics came up, which demanded for accurate information about these topics. Search engines are not capable to satisfy these needs and led to the construction of domain specific search engines.

Domain specific search engines attempt to collect pages that are relevant to a particular domain and them to their index. In order to achieve this goal, they utilize focused crawlers, which are agents that utilize the graph of the Web to find pages or documents that belong to a particular topic. Additionally, focused crawlers are used for the construction of thematic web portals and their maintenance with new and updated information.

A variety of methodologies have been proposed for the aforementioned problem. In [25] the crawler is based on link criteria whereas in [26] the system exploits the knowledge about domains in order to construct topic hierarchies. Other approaches [27, 28, 29], [30, 31] adopt learning and evolutionary algorithms respectively. In our work we frame the problem of focused crawling using Reinforcement Learning to learn an optimal crawling strategy.

4.1 Reinforcement Learning

Reinforcement Learning (RL) addresses the problem of how an agent can learn a behaviour through trial-and-error interactions with a dynamic environment [32]. In an RL task the agent, at each time step, senses the environment's state, s_t in S , where S is the finite set of possible states, and selects an action a_t in $A(s_t)$ to execute, where $A(s_t)$ is the finite set of possible actions in state s_t . The agent receives a reward, r_{t+1} in R , and moves to a new state s_{t+1} . The objective of the agent is to maximize

the cumulative reward received over time. More specifically, the agent selects actions that maximize the expected discounted return:

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$$

where γ is the discount factor and expresses the importance of future rewards. A *policy* π specifies that in state s the probability of taking an action a is $\pi(s,a)$. For any policy π the *action-value function*, $Q^\pi(s,a)$, can be defined as the expected discounted return for executing a in state s and thereafter following π . The optimal policy, π^* , is the one that maximizes the action-value $Q^\pi(s,a)$ for all state-action pairs. In order to learn the optimal policy, the agent learns the optimal action-value function, Q^* which is defined as the expected return of taking action a in state s and thereafter following the optimal policy π^* . The most widely used algorithm for finding the optimal policy is the Q-learning algorithm [33].

4.2 Methodology

To formulate the problem of focused crawling as an RL problem we must define the following components:

A set of states, S : The state is defined as the pages accessed by the crawler, as the perception of the environment arises mainly by the pages retrieved at any given time.

A set of actions, A : Actions are defined as the categories of links, among which the crawler has to choose when visiting a specific page.

A reward function, $r(s,a)$: The crawler receives reward r when it visits page, s :

$$r = \begin{cases} 1 & \text{if } s = \text{target page} \\ 0 & \text{otherwise} \end{cases}$$

In the training phase the crawler executes a number of episodes starting from a seed set of pages. The episode ends when the crawler finds a relevant page or reaches a number of predefined steps. When the crawler visits a page it classifies the hyperlinks that the page contains, based on the textual information of the page that they link to. The crawler chooses a category of links from where it will select a URL to follow. The crawler receives a reward and transits to the next page. In order to explore the state space we make use of ϵ -greedy action selection method, where an action a is selected according to the following rule:

$$a = \begin{cases} a \text{ random action with probability } \epsilon \\ \arg \max_{a'} Q(s,a') \text{ with probability } 1 - \epsilon \end{cases}$$

In the crawling phase the crawler starts with a set of seed pages and uses a queue where it maintains a list of unvisited hyperlinks. The hyperlinks from each page are extracted and classified. Each action is evaluated using the function that crawler learned in the training phase. The crawler selects the link with the highest relevance score and repeats the same procedure until it visits a certain number of pages or when the queue is empty.

In order to evaluate the proposed methodology, the sport of snowboarding has been chosen as the specific topic for conducting the experiments. A dataset was collected by a breadth-first search crawl starting from the Winter Sports web page of dmoz¹. We compared our approach with the widely used Best First strategy where the best hyperlink according to a criterion is selected and the page that it links to is fetched. We also implemented a variant of our approach with a different definition of the actions.

¹ http://dmoz.org/Sports/Winter_Sports/

The results have shown that the proposed method approaches the behavior of the Best First strategy and outperforms the variant of the reinforcement learning approach. The results encourage us to improve the proposed method and also identify the types of task for which a learning agent might be better than the best-first heuristic.

5. Conclusions

Because of the special characteristics of the new kind of data that are nowadays available (e.g. biological data), the variety of new problems and the extremely high importance of machine learning research, a large number of critical issues is still open and demands active and collaborative research by the academia as well as the industry. Moreover, new technologies led to a constantly increasing number of new questions on new data. The scientific community demands and machine learning provides the opportunities for novel and improved methods for analyzing all these data.

In this paper, some emerging trends of machine learning have been presented. Although mining from Web and biological data are considered as significantly upcoming topics in the field, there are also other interesting topics that were not discussed in this paper. Some of them are applications on MRI data, astronomical data, robotics, video games, music data etc. Other recent trends in the research of machine learning include learning from spatial and visual data. Machine learning can contribute to a better understanding of human interpretation and recognition of real world scenes, as well as of improving the capability of artificial vision systems. The most used learning method in computer vision problems is supervised learning and also Reinforcement Learning in some robotic vision tasks.

References

- 1 Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Washington, DC, USA, May 26-28, 1993, 207-216.
- 2 Agrawal R, Srikant R. Mining sequential patterns. In *Proceedings of the 11th IEEE ICDE International Conference on Data Engineering*, pages 3-14, 1995.
- 3 Ma Q, Wang J T L. Biological data mining using Bayesian neural networks: A case study. *International Journal on Artificial Intelligence Tools*, Special Issue on Biocomputing, 1999, 8(4), 433-451.
- 4 Hirsh H, Noordewier M. Using background knowledge to improve inductive learning of DNA sequences. *Proceedings of the 10th IEEE Conference on Artificial Intelligence for Applications*, 1994, 351-357.
- 5 Zien A, Ratsch G, Mika S, Scholkopf B, Lengauer T, Muller R-K. Engineering support vector machine kernels that recognize translation initiation sites. *Bioinformatics*, 2000, 16(9), 799-807.
- 6 Houle J L, Cadigan W, Henry S, Pinnamaneni A, Lundahl S. Database mining in the human genome initiative. Whitepaper, Bio-databases.com, Amita Corporation., March 10th 2004, Available: <http://www.biodatabases.com/whitepaper.html>
- 7 Li J, Ng K-S, Wong L. Bioinformatics adventures in database research. *Proceedings of the 9th International Conference on Database Theory*, 2003, Siena, Italy, 31-46.
- 8 Zeng F, Yap C H R, Wong L. Using feature generation and feature selection for accurate prediction of translation initiation sites. *Genome Informatics*, 2002, 13, 192-200.
- 9 Kozak M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research*, 1987, 15(20) 8125-8148.
- 10 Tzanis G, Berberidis C, Alexandridou A, Vlahavas I. Improving the accuracy of classifiers for the prediction of translation initiation sites in genomic sequences, In *Proceedings of the 10th Panhellenic Conference on Informatics*, Bozanis P, Houstis E N (Eds.), Springer-Verlag, LNCS 3746, 426-436, Volos, Greece, 11-13 November, 2005.
- 11 Tzanis G, Vlahavas I. Prediction of translation initiation sites using classifier selection. In *Proceedings of the 4th Hellenic Conference on Artificial Intelligence*, Antoniou G, Potamias G,

- Plexousakis D, Spyropoulos C (Eds.), Springer-Verlag, LNAI 3955, 367-377, Heraklion, Crete, 18-20 May, 2006.
- 12 Pedersen A G, Nielsen H. Neural network prediction of translation initiation sites in eukaryotes: Perspectives for EST and genome analysis. In Proceedings of the 5th International Conference on Intelligent Systems for Molecular Biology, AAAI Press, Menlo Park, California, USA (1997) 226-233.
 - 13 Katakis I, Tsoumakas G, Vlahavas I. Email mining: Emerging techniques for email management (to appear), Web Data Management Practices: Emerging Techniques and Technologies, Athena Vakali, George Pallis (Ed.), Idea Group Publishing, 32, 2006.
 - 14 Bickel S, Scheffer T. Learning from message pairs for automatic email answering. In Proceedings of the 15th European Conference on Machine Learning, Pisa, Italy, 2004.
 - 15 Busemann S, Schmeier S, Arens R G. Message classification in the call center In Proceedings of the 6th Conference on Applied Natural Language Processing, Seattle, Washington, 2000.
 - 16 Clarck J, Koprinska I, Poon J. A neural network based approach to automated e-mail classification. In Proceedings of the IEEE/WIC International Conference on Web Intelligence, 2003.
 - 17 Klimt B, Yang Y. The Enron corpus: A new dataset for email classification research. In Proceedings of the 15th European Conference on Machine Learning, Pisa, Italy, 2004.
 - 18 Graham-Cumming J. PopFile: Automatic email classification (<http://popfile.sourceforge.net>), 2002.
 - 19 Ho V. EMMA: An e-mail management assistant. In Proceedings of the 2003 IEEE/WIC International Conference on Intelligent Agent Technology (IAT 2003), 13-17 October 2003, Halifax, Canada.
 - 20 Muresan S, Tzoukerman E, Klavans, J L. Combining linguistic and machine learning techniques for email summarization. In Proceedings of CoNLL-2001, Toulouse, France, 2001.
 - 21 Sahami M, Dumais S, Heckerman D, Horvitz E. A Bayesian approach to filtering junk e-mail. In Learning for Text Categorization: Papers from the 1998 Workshop. Madison, Wisconsin: AAAI Technical Report WS-98-05, 1998.
 - 22 Drucker H, Vapnik V, We D. Support vector machines for spam categorization. IEEE Transactions on Neural Networks, 1999, 10(5), 1048-1054.
 - 23 Sakkis G, Androutsopoulos I, Paliouras G, Karkaletsis V, Spyropoulos C D, Stamatopoulos P. A memory-based approach to anti-spam filtering for mailing lists. Information Retrieval, 2003, 6(1), 49-73.
 - 24 Katakis I, Tsoumakas G, Vlahavas I. On the utility of incremental feature selection for the classification of textual data streams, In Proceedings of the 10th Panhellenic Conference on Informatics, Bozanis P, Houstis E N (Eds.), Springer-Verlag, LNCS 3746, 338-348, Volos, Greece, 11-13 November, 2005.
 - 25 Cho J., Garcia-Molina H, Page L. Efficient crawling through URL ordering. Computer Networks and ISDN Systems, 1998, 30(1-7) 161-172.
 - 26 Chakrabarti S, Van den Berg M, Dom B. Focused crawling: a new approach to topic specific Web resource discovery. Computer Networks, 1999, 31(11-16) 1623-1640.
 - 27 O'Meara T, Patel A. A topic specific web robot model based on restless bandits. IEEE Internet Computing, 2001, 5(2) 27-35.
 - 28 Rennie J, McCallum A. Efficient web spidering with reinforcement learning. In Proceedings of the 16th International Conference on Machine Learning, 1999, 335-343.
 - 29 Grigoriadis A, Paliouras G. Focused crawling using temporal difference learning. In Proceedings of the 3rd Hellenic Conference on Artificial Intelligence, 2004, 142-153.
 - 30 Johnson J, Tsioutsoulouklis K, Giles C L. Evolving strategies for focused web crawling. In Proceedings of the 20th International Conference on Machine Learning.
 - 31 Pant G, Menczer F. MySpiders: evolve your own intelligent web crawlers. Autonomous agents and multi-agent systems, 2002, 5(2) 221-229.
 - 32 Sutton R S, Barto A G. Reinforcement Learning, An Introduction. MIT Press, 1999.
 - 33 Watkins C, Dayan P. Q-learning. Machine Learning, 1992, 8, 279-292.

An Analytical Strategic Model for the Configuration of a Closed-Loop Supply Chain Network

Anastasios Xanthopoulos, Eleftherios Iakovou

Aristotle University of Thessaloniki, Dept. of Mechanical Engineering, Industrial Management Division, Laboratory of Quantitative Analysis, Logistics and Supply Chain Management, University Campus, 54124 Thessaloniki, P.O. Box 461, Greece, {axanthop, eiakovou}@auth.gr

Financial, environmental and legislative motives impose today the optimal exploitation of the electronic, electrical and mechanical products that reach the end of their useful life. This situation has as a result enhanced the need for a joint and holistic treatment of the forward supply channel together with the reverse flow of used products. Thus, the complexity of production planning in such manufacturing/industrial environments is significantly increased. Moreover, in this framework the configuration of a concrete and optimal closed-loop supply chain network gains remarkable importance. More particularly, the main objective of this work is to present a framework for the development of an analytical strategic mixed integer linear programming model, which will provide support in the relevant decision-making processes. This model will aim at maximising the profit in a finite time horizon of multiple periods, while it should take, in parallel, into account the legislative and environmental restrictions that are placed for the under review product types.

Keywords

Closed-loop supply chain, decision-making tool framework, recovery, used products.

1. Introduction

The past two decades constituted a landmark with regard to the perception of the exclusive traditional forward character of supply chain management. The over-consumption and the ever shorter products' lifecycles had as a consequence the accumulation of enormous volumes of waste. Electronic, electrical and mechanical products possess the lion's share in this reality. This situation had as a result the emergence of the need to exploit optimally the products that reach the end of their useful life. The establishment of Community and National laws and other regulatory interventions for the protection of environment constituted a direct outcome of the new reality. On the other hand, this occasion gave the chance to the manufacturing industry to realize that the recovery of end-of-life products can incorporate a great potential as an important source of profit/value. However, the need for a joint and holistic treatment of the forward supply channel together with the reverse flow and recovery of used products increases significantly the complexity of production planning in such manufacturing/industrial environment. Thus, the configuration of a concrete and optimal closed-loop supply chain network (per case) gains remarkable importance.

In this framework a great number of issues arises that must be optimally tackled. For instance, we should investigate whether the under development network will lead in the significant transformation of the traditional forward supply network, and whether this reverse network will be exclusively privately structured or we will make use of subcontracting and in what degree. The decisions regarding the type, location, number and capacity of the collection, sorting, warehousing, disassembling, and recovering installations are in the heart of the problem. Furthermore, it is very important that we determine the optimal agreement/contract with the third-party reverse logistics companies, if we finally decide to proceed to such a strategic decision.

This work presents a decision-making tool for the development of closed-loop supply chain networks, especially for the electronic, electrical and mechanical products' industry. More specifically, in this

paper we present as a first step the framework of developing an analytical strategic mixed integer linear programming model, so as to tackle the investigated problem. The under development model, as a point of differentiation from the previous efforts will give a more generalized form of the problem; something that will also contribute to the maximisation of its usefulness. In any case, this model will have as its main objective the maximisation of the net present value of the profit in a finite time horizon of multiple periods, while it would also take into account the legislative and environmental restrictions that are placed for the under review product types.

The rest of the paper is organised as follows: the second section presents an up-to-date literature review of the scientific work related to quantitative methodologies for the optimal configuration of closed-loop supply chains. In the third section, we elaborate on the investigating problem and we give the major model assumptions to be considered. The next section describes the objective function and constraints of the under development mathematical problem, along with its possible extensions. Following, the fifth section refers to the necessary sensitivity analysis and to the most important scenarios that should be examined from a strategic point-of-view. Finally, we conclude by summarizing the value of this work and provide some research directions for the future.

2. Literature review

The majority of the past research efforts regarding the optimal configuration of closed-loop supply chain networks involved the use of mathematical programming. The first research step had to do with the appropriate modification of the corresponding location models from the forward supply chain problem. One of the most important works in this field is that of Fleischmann [1]. Fleischmann formulated a single-period mixed integer-programming model so as to address the investigating problem. In this setting, the model considers three levels of facilities for a single type of product, namely test centres, factories, and distribution warehouses. The main scope of this work is to minimise the total costs and at the same time to take into account the co-ordination issues between the forward and the reverse supply channel.

With regard to the research efforts addressing the open-loop reverse logistics network design problem, Caruso et al. [2] developed a quantitative location model for a waste management system. This model determines the number and the location of waste disposal plants, specifying the technology adopted, the amount of waste processed and the service basin of each plant. Afterwards, Beamon and Fernandes [3] developed a multi-period mixed integer programming model, so as to determine which warehouses and collection centres should be opened, which warehouses should have sorting capabilities and how much material should be transported between each pair of sites. The main goal of this model is the minimisation of the net present value of the required investment and operational costs. We should also mention the remarkable work of Jayaraman et al. [4] as well as this of Marin and Pelegrin [5], who developed mathematical programming models for the cost minimisation of the reverse network configuration problem.

At the same time, a considerable number of research works exist that focuses on specific corporate and business problems. Kroon and Vrijens [6] developed a single-period, cost minimization mixed integer linear programming model for designing a return logistics system for returnable containers for a large logistics service organization in The Netherlands. The model takes into account the transportation, maintenance and storage processes of the returned products, along with the decisions concerning the number and the location of the containers' warehouses. Similarly, Barros et al. [7] developed an analogous cost minimisation methodology for the configuration of reverse logistics and recycling networks. Afterwards, Spengler et al. [8] developed a similar approach for the optimisation of the recycling network of the industrial by-products of German steel industry. Following, Krikke et al. [9] developed a mixed linear programming model in order to optimise the remanufacturing network of Océ; a leading producer of copying, printing and plotting systems with headquarters in The Netherlands. Finally, Bloemhof-Ruwaard et al. [10] in a more recent work developed noteworthy single-period mixed integer linear programming models for the recovery network configuration for the cases of the refrigerators and the waste paper.

3. Problem statement

In the past few years, the design and configuration of reverse logistics supply chains is an evergreen and at the same time very important issue for both Greek and international industry. The efficient management of returned products' flows, in such a way that the profitability is increased and the synergies with the traditional forward chain are handled properly, constitutes a major challenge for the closed loop supply network configuration problem. In this work, we treat this particular problem via the presentation of a generic framework for the development of multi-period, multi-product, mixed integer linear programming models that address the optimal decision-making processes of the investigating problem. More specifically, we propose the development of a generalised strategic model for configuring closed loop supply chains, while aiming at the maximisation of profit and meeting the legislative and environmental constraints that are placed.

Figure 1, depicts the decision-tree of the under review discrete scenarios of developing a closed loop supply network. The development of such a network presupposes the existence of the traditional supply chain, which is to be suitably transformed so that an effective closed loop network will finally result. Of course, this case concerns the total of the present industrial enterprises that have already developed their forward supply network and they are to be involved in product recovery processes. Starting-point in any case is the end-of-life products' returns. Thereafter, we have three alternative choices: a) to contract with the state collective system for the management of the waste electronic, electrical and mechanical equipment, provided that such a system exists in each under review country; in Greece for instance there exists the Appliances Recycling S.A. which undertakes the management of the electronic and electrical waste for a price proportional to the type and the weight of the collected products, b) to co-operate with third-party reverse logistics providers that will undertake for a pre-agreed cost the management of the necessary recovery processes and will finally return to the producers the value recovered from the end-of-life products, and c) to develop/acquire privately-owned installations/equipment for the collection, sorting, storage and distribution of the returned products. These private collection and warehouse installations can be located and integrated in the same facility or else independently but in the same geographical region. Moreover, their development can be associated with the appropriate adaptation or expansion of the existing forward channel facilities or they can constitute a part of a discrete reverse channel. Provided that we come to a decision to develop our own private collection and storage facilities, then we can again proceed to an agreement with third-party reverse logistics providers but from the point of the disassembly process and on. Alternatively, we can develop privately-owned disassembly facilities. Supplementary to the two aforementioned choices it is likely also to develop refurbishing installations, in order to process properly the returned products and channel them to the secondary market for reuse; again these facilities can result from the proper adaptation or expansion of the existing forward supply facilities or they can be a part of a distinct reverse channel. We should also mention here that a part from the collected and sorted products it is likely to be disposed of because of its bad condition and unsuitability for any recovery process. Afterwards, by considering the path of the private disassembly facilities, we can then channel the disassembled end-of-life products to private recycling and remanufacturing installations, or else to collaborate with third-party reverse logistics providers for the final recovery processes. Finally, some of the disassembled products will be rejected due to their low quality and inappropriateness for any recovery operation.

Below, we give the most important assumptions that must be taken into account in the proposed model:

- The recovered products – recycled and/or remanufactured – can satisfy only a specific part of the demand in new products, components and materials. This realistic assumption holds so as to avoid the uncertainty in production planning and scheduling and the variability in a reverse logistics network, i.e. unknown timing, quantity and quality of product returns. By considering such a policy, it is very possible that we increase the efficiency of the system and balance the productive and operational processes, while limiting our exposure to the innate uncertainties of reverse logistics.

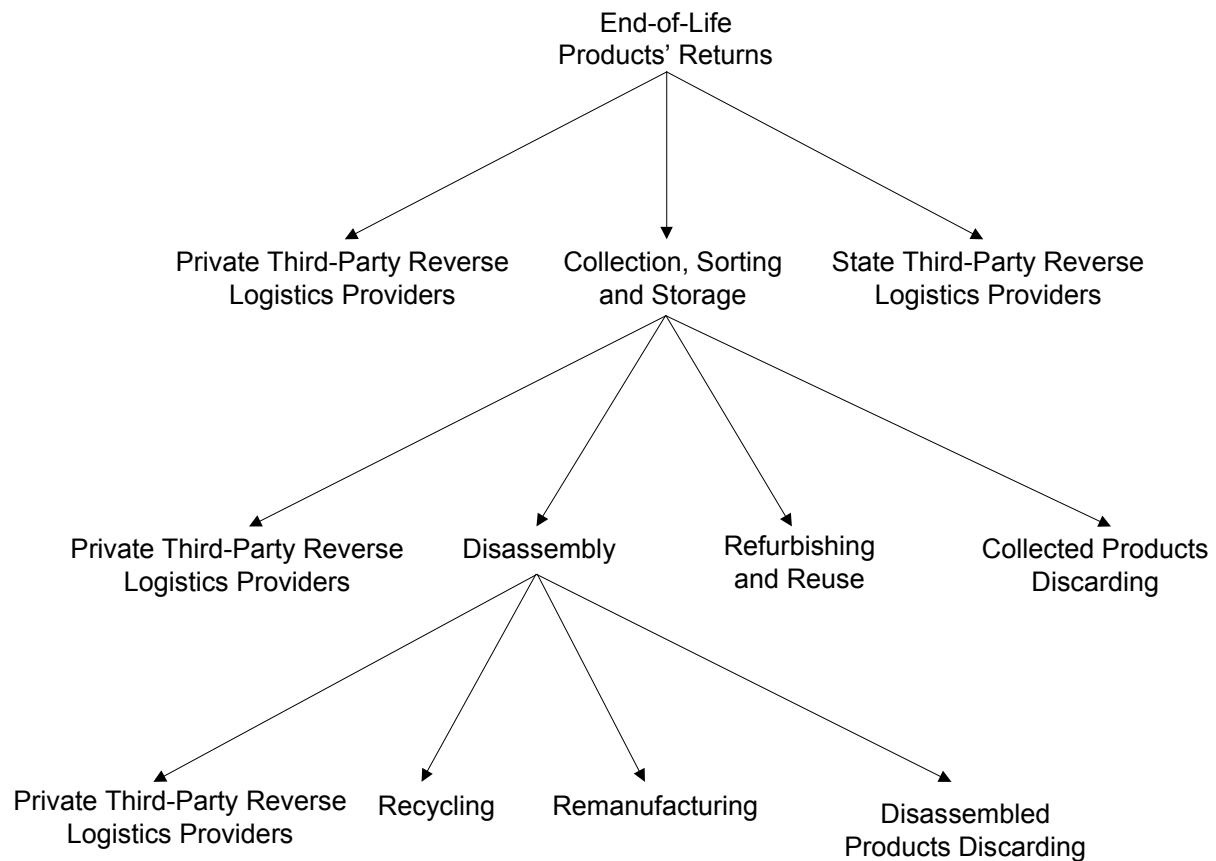


Figure 1 Alternative Scenarios of Closed-Loop Supply Chain Network Development

- In many cases, a single uniform policy is preferred for the total of the under review types of products. In such a way, we limit the complexity of the under development network and we drastically improve the control and monitoring over it.
- The corporate revenues can be increased through the development of an ecological/green profile, due to marketing reasons. A contract with the state third-party reverse logistics providers is an example towards this direction.
- Lead-times and machines' set times should not be taken into consideration in the modelling if they are shorter from the unit time-step (period) of the model.
- The duration of the agreement with the third-party reverse logistics providers will be as much as this of the examining time horizon.
- In case of contracting with a private third-party company we should have the right to decide which percent of the disassembled products will be recovered through recycling and which through remanufacturing.

It should be mentioned here that the warehouse and distribution centres are selected to be before the disassembly process' stage, because the holding and capital cost is bigger for a disassembled product than for a collected but not disassembled one. Finally, some more factors that we should consider in the development of the quantitative model are the following: the choice not to collect a part of the available returned products, the delayed satisfaction of demand (backorders), the case of lost sales, as well as the option to dispose of to a landfill the collected products that cannot be recovered in any way because of their disrepair state (bad condition/quality).

4. Framework for the quantitative model formulation

In this section, we present a framework for the development of a generic strategic mixed integer linear programming model, so as to support the decision-making processes regarding the configuration of a closed-loop supply chain network.

4.1 Nomenclature

Below, we provide the definition of the necessary indicators that should be used:

- $a \in \{1, \dots, A\}$: the alternative options to contract with private third-party reverse logistics providers for the total of the reverse logistics activities.
- $b \in \{1, \dots, B\}$: the alternative options to contract with private third-party reverse logistics providers for the stage of the disassembly process and on.
- $c \in \{1, \dots, C\}$: the alternative options to contract with private third-party reverse logistics providers only for the remanufacturing and recycling processes.
- $d \in \{1, \dots, D\}$: the alternative scenarios to develop disassembly facilities as regards their number, capacity, location and type.
- $i \in \{1, \dots, I\}$: the under review products' types.
- $j \in \{1, \dots, J\}$: the alternative scenarios to develop collection, sorting, warehouse and storage facilities as regards their number, capacity, location and type. This set of decisions can be divided into three distinct subsets: $\{1, \dots, J_1, J_1+1, \dots, J_2, J_2+1, \dots, J\}$. The first and the second subset concern the case of the proper adaptation and expansion, respectively, of the existing forward network, and the third one the case of developing a discrete reverse logistics network.
- $r \in \{1, \dots, R\}$: the alternative scenarios to develop refurbishing for products' reuse facilities as regards their number, capacity, location and type. This set of decisions can be divided again into three distinct subsets: $\{1, \dots, R_1, R_1+1, \dots, R_2, R_2+1, \dots, R\}$. The first and the second subset concern the case of the proper adaptation and expansion, respectively, of the existing forward network, and the third one the case of developing a discrete reverse logistics network.
- $t \in \{1, \dots, T\}$: the number of the discrete periods in the examining time horizon.
- $u \in \{1, \dots, U\}$: the alternative scenarios to develop remanufacturing facilities as regards their number, capacity, location and type.
- $v \in \{1, \dots, V\}$: the alternative scenarios to develop recycling facilities as regards their number, capacity, location and type.

Moreover, as IR we define the interest rate that will be used for the present value calculation of the profit.

4.2 Model formulation

Initially, we describe the objective function:

Max $\sum_i \sum_t$ (Value that derives from the agreement with the state reverse logistics service providers (green profile) + Revenues from the reused and remanufactured products and components, as well as from the recycled materials) $_{it} \cdot (1 + IR)^{-t} - \sum_i \sum_t$ (Contribution to the state collective system for waste management services) $_{it} - \sum_a$ (Fixed cost of agreement with a private third-party reverse logistics provider for the total of the reverse logistics processes) $_a - \sum_b$ (Fixed cost of agreement with a private

third-party reverse logistics provider for the stage of disassembly process and on) $_{ib} - \sum_c$ (Fixed cost of agreement with a private third-party reverse logistics provider for the remanufacturing and recycling processes) $_c - \sum_i \sum_t [\sum_a$ (Variable cost per product that stems from the agreement with a private third-party reverse logistics provider for the total of the reverse logistics processes) $_{ita} + \sum_b$ (Variable cost per product that stems from the agreement with a private third-party reverse logistics provider for the stage of disassembly process and on) $_{itb} + \sum_c$ (Variable cost per product that stems from the agreement with a private third-party reverse logistics provider for the remanufacturing and recycling processes) $_{itc}] \cdot (1 + IR)^{-t} - \sum_j$ (Fixed cost of developing collection, sorting, warehouse and distribution facilities) $_j - \sum_i \sum_t \sum_j$ (Collection, sorting, holding and distribution cost per product) $_{ij} \cdot (1 + IR)^{-t} - \sum_j \sum_r$ (Fixed cost of developing refurbishing facilities) $_{jr} - \sum_i \sum_t \sum_j \sum_r$ (Transportation and refurbishing cost per product) $_{ijr} \cdot (1 + IR)^{-t} - \sum_j \sum_d$ (Fixed cost of developing disassembly facilities) $_{jd} - \sum_i \sum_t \sum_j \sum_d$ (Transportation and disassembly cost per product) $_{ijd} \cdot (1 + IR)^{-t} - \sum_d \sum_u$ (Fixed cost of developing remanufacturing facilities) $_{du} - \sum_i \sum_t \sum_d \sum_u$ (Transportation and remanufacturing cost per product) $_{idu} \cdot (1 + IR)^{-t} - \sum_d \sum_v$ (Fixed cost of developing recycling facilities) $_{dv} - \sum_i \sum_t \sum_d \sum_v$ (Transportation and recycling cost per product) $_{itdv} \cdot (1 + IR)^{-t} - \sum_i \sum_t$ (Penalty cost for not collecting the available returned products + Cost for disposing of properly the end-of-life products (either disassembled or not) + Cost of backorders + Cost of not satisfying the demand in remanufactured and recycled products) $_{it} \cdot (1 + IR)^{-t}$.

Thus, the objective function aims at optimizing the closed-loop supply chain network via the maximisation of the revenues from the end-of-life products' recovery minus the investment, tactical and operational costs. To this effect, we should also develop multiple groups of constraints, which we describe right below:

- Capacity constraints for the under development collection, warehouse, disassembly, refurbishing, remanufacturing and recycling facilities.
- Upper limits for the end-of-life products that the third-party reverse logistics providers will process. These limits are case-specific and depend on the type of the pre-specified commitment.
- Constraints that ensure that a single policy will be applied to the under review types of products.
- Inequalities that ensure the implementation of a uniform policy for the collection, warehouse and refurbishing facilities.
- Continuity equations from the point of the products' return and collection to the satisfaction of the demand in remanufactured and recycled products.
- Initial values for the inventory position of each product.
- Initial and ending values for the backorders of each product.
- Environmental constraints that place an upper limit to the uncollected end-of-life products.
- Legislative constraints that impose a lower limit to the (weight) percentage of the end-of-life products that will be finally recovered. For example, we can mention here the European Union Directive for the management of the Waste Electrical and Electronic Equipment (WEEE).

- Constraints that give an upper acceptable limit for the unsatisfied demand in remanufactured and recycled products.
- Inequalities that indicate that some of the end-of-life products will be of such bad condition/quality for reuse and/or for the remanufacturing process.
- Upper and lower limits for the returned products to be disposed of. A stochastic part of the end-of-life products will be of low quality for any value recovery process, while some of these products can be selectively recovered in order to meet a legislative or environmental constraint.
- The variables that are related to the development of the network facilities and to the agreements with third-party providers are binary. The remainder variables can be either non-negative continuous or general integers.

Full details for the mathematic formulation of the model can be found in the working paper of Xanthopoulos and Iakovou [11].

4.3 Model extensions

The proposed model can be extended by relaxing the single-policy constraint, so as to allow each type of product to be treated in a different way. In other words, the optimal solution may indicate that some of the end-of-life products will pass through the privately-owned reverse logistics network, while the rest of them via the channels of the state and private third-party reverse logistics providers. Moreover, we can take into consideration the possibility that the recovery technology will be changed dramatically in the next few years. Finally, the possibility that the end-of-life products' returns will be significantly increased in the near future constitutes another extension.

5. What-if analysis

The main goal of this paper is to present a decision-making tool for strategic management decisions. To this effect, we presented in the previous sections a framework for the development of a mathematical model for a specific problem. However, for the sake of completeness the model should be enriched with the necessary sensitivity analysis by investigating the most interesting what-if scenarios. The innate uncertainty and variability in a closed-loop supply chain 'highlight' how significant is the what-if analysis in supporting the relevant decision-making processes. Following, we provide some essential directions for a concrete analysis:

- In what way do changes in the unit variable costs of recovering a product influence the optimal development of the reverse logistics network? Are the agreements with the third-party providers sensitive to such changes?
- Is there any direct correlation between the unit variable costs of the model with how easy is to meet the legislative and environmental restrictions?
- In case that we have a significant reduction in the products' returns, does the absence of economies of scale lead us to an agreement with the state collective system of waste management? Similarly, what effect might have on the final decisions a considerable increase in products' returns?
- If some of the under development recovery facilities are located in the same place, does this scenario constitute an attractive option? In such a case the internal transportation costs will be decreased.
- What effect has the severity of the legislative constraints on the final decisions?
- What is the impact of the interest rate on the decision-making processes?
- What are the consequences from a considerable change in the unit discarding cost of the low quality end-of-life products in their recovery ratio?

6. Conclusions

In this paper we presented a decision-making tool framework for the optimisation of configuring a closed-loop supply chain network. More specifically, we described the development of a generic model, which can be used by corporations that have already structured their forward supply chains and are ready to embark on recovery processes. This model as a point of differentiation from the previous efforts presents the investigating problem in its generalized form; something that enhances its usefulness. For instance, we included the co-operation issues with state and private third-party reverse logistic providers, as well as the development of all the necessary recovery facilities for an integrated private network. Finally, the main direction of our future research will be the development of a holistic model that will treat jointly the forward and reverse supply chain network configuration for newly-established companies and the network reconfiguration problem for the existing ones. Moreover, it is our intention to investigate the pure tactical and operational issues of the problem.

References

- 1 Fleischmann M. Quantitative Models for Reverse Logistics. PhD Thesis, Erasmus University Rotterdam, 2000.
- 2 Caruso C, Colorni A, Paruccini M. The Regional Urban Solid Waste Management System: A Modelling Approach. *European Journal of Operational Research* 1993; 70: 16-30.
- 3 Beamon B M, Fernandes C. Supply Chain Network Configuration for Product Recovery. *Production Planning and Control* 2004; 15(3): 270-281.
- 4 Jayaraman V, Patterson R A, Rolland E. The Design of Reverse distribution Networks: Models and Solution Procedures. *European Journal of Operational Research* 2003; 150: 128-149.
- 5 Marin A, Pelegrin B. The Return Plant Location Problem: Modelling and Resolution. *European Journal of Operational Research* 1998; 104: 375-392.
- 6 Kroon L, Vrijens G. Returnable Containers: An Example of Reverse Logistics. *International Journal of Physical Distribution & Logistics Management* 1995; 25(2): 56-68.
- 7 Barros A I, Dekker R, Scholten V. A Two-Level Network for Recycling Sand: A Case Study. *European Journal of Operational Research* 1998; 110: 199-214.
- 8 Spengler T, Püchert H, Penkuhn T, Rentz O. Environmental Integrated Production and Recycling Management. *European Journal of Operational Research* 1997; 97: 308-326.
- 9 Krikke H R, Van Harten A, Schuur P C. Business Case Océ, Reverse Logistics Network Redesign for Copiers. *OR-Spektrum* 2001; 21(3): 381-409.
- 10 Bloemhof-Ruwaard J M, Krikke H, Van Wassenhove L N. OR Model for Eco-Eco Closed-Loop Supply Chain Optimization, in *Reverse Logistics: Quantitative Models for Closed-Loop Supply Chains*. Springer Press, Berlin, 2004.
- 11 Xanthopoulos A, Iakovou E. A Generic Analytical Quantitative Model for the Design of a Reverse Supply Chain. Working Paper, Aristotle University of Thessaloniki 2006.

SESSION 3:

**POLITICAL AND
SOCIAL ISSUES**

A Theoretical Analysis of the Europeanisation Impact in Macedonia

Ismail Aydil¹, Andrew J. Taylor², Miltiadis Sarigiannidis³

¹SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, iaydil@seerc.info

²University of Sheffield, Dept. of Politics, Elmfield Northumberland Road Sheffield, S10 2TU UK
a.j.taylor@sheffield.ac.uk

³City College, Aristotle University of Thessaloniki, Dept. of Law, Thessaloniki, Greece,
msarigiannidis@seerc.info

The increasing need to understand the international, specifically European, involvement in South East Europe created confusion on the theoretical framework for analysis. Since the countries of the region are not yet involved in the European integration process at the same time considered associated countries of all the European institutions, it is an important challenge to pinpoint the process of europeanisation for the South East European countries. In this paper, I try to analyse recent political developments in Former Yugoslav Republic of Macedonia with regard to Europeanisation. I assert that the europeanisation process in Macedonia did not start with its official acceptance as a candidate state to the European Union in early 2006, but has been a long and painful journey. In the first part, I present an analysis of the definition of Europeanisation. Since it is not the main purpose, it does neither intend to cover all of the europeanisation literature nor comprehensively analyse it. Second, I present the affect of Europeanisation on 2004 enlargement of the Union. According to this analysis, third, I come into the details of the impact of europeanisation in South East Europe. Fourth, I summarise the recent political developments in FYR Macedonia and conclude with an analysis of europeanisation impact in the country.

Keywords

Decentralisation, europeanisation, Macedonia, multi-level governance, South East Europe.

1. Defining Europeanisation

There is a consensus among the scholars on the nature of the variables involved in europeanisation, namely the EU institutions and member (+ candidate) states. We should keep in mind that europeanisation does not mean European integration and it does not intend to change European integration studies into a new theoretical framework. European integration studies had been focusing on explaining the political phenomena at the European level. Rather, europeanisation focuses on the domestic level [1]. It is related to the study of European integration but it is different from it.

Europeanisation is always (to a certain extent) a process of domestic political change caused (somehow) by processes of European integration. (p. 8, [2])

In a different way, it can be said that there are certain points left unexplained by the European integration studies [3]. In a more explanatory frame, it is contended that Europeanization is about

[h]ow do European Integration and Europeanization more generally affect domestic policies, politics and polities of the member states and beyond. (p. 1, [4])

The need to analyse the domestic policy making under the influence of European involvement is relevant when there is an increasing observation of changing institutional mechanisms. Europeanisation is

the emergence and development at the European level of distinct structures of governance, that is, of political, legal and social institutions associated with political problem solving that formalizes interactions among the actors, and of policy networks specializing in the creation of authoritative European rules. [emphasis original] (p. 3, [5])

These new governance structures have been on the agenda of the research community for some time. What the concept of europeanisation attempts to do is to search for an analytical pattern that can be applied to further observation.

The influence of the EU however does appear outside the confines of the member states as well. It is becoming more important in this sense to resort to the europeanisation literature in analysing the European influence on candidate countries as well as associated countries at various levels.

Europeanisation has been employed to refer to examples where distinct European forms of organisation and governance have been exported outside Europe's territorial boundaries. (p. 14, [1])

The europeanisation in this sense focuses on the policy spheres of domestic policy making. Europeanisation involves changes in external boundaries of states with the involvement of European institutions [6]. This way of analysis apparently refers to the enlargement process. A wider definition that embraces various levels of European involvement asserts that europeanisation is:

[T]he development and sustaining of systematic European arrangements to manage cross-border connections, such that a European dimension becomes an embedded feature which frames politics and policy within the European states. (p. 370, [7])

With variations involved, some defined europeanisation as understanding how Europe matters in the analysis of the domestic polity [1]. In this type of approach the cross-country and cross-sector evaluations become the basis of research. Accordingly, for some scholars the organisational analysis becomes the central tenet.

Europeanisation is an incremental process reorienting the direction and shape of politics to the degree that EC political and economic dynamics become part of the organisational logic of national politics and policy-making (p. 17, [8]).

A similar idea might be employed with a variation of focusing on the policy areas under European influence. In this sense, Europeanisation is

a process by which domestic policy areas become increasingly subject to European policy-making. (p. 574, [9])

In the final analysis, Radaelli comes up with a more comprehensive definition in this train of thought:

Processes of (a) construction (b) diffusion and (c) institutionalisation of formal and informal rules, procedures, policy paradigms, styles, 'ways of doing things' and shared beliefs and norms which are first defined and consolidated in the making of EU decisions and then incorporated in the logic of domestic discourse, identities, political structures and public policies (p. 30, [10]).

This definition is well structured along the lines of constructivist – sociological – institutionalism. Employing this wide definition of europeanisation directs towards a constructivist analysis of Multi-Level Governance.

An increasing amount of literature arising both from historical institutionalism and from rational institutionalism appear around the concept of Multi-Level Governance (MLG). The MLG approach in Europeanisation is particularly strong in the analysis of the interaction between sub-national and supra-national levels of governance in the emerging European polity. First, the emphasis of MLG is on the actors other than state institutions in the making of policy. In a way MLG focuses on the interplay of public and private actors in the policy making [11]. Second, the important contribution of MLG to europeanisation is the reliance on different levels of policy making, the sub-national, national, and supra national levels of interaction among various types of actors and policy makers [12]. Third, MLG indicates the erosion of hierarchical governance mechanisms and the increasing amount of hybrid governance arrangements with emphasis on horizontal in addition to vertical inter-organizational

architectures [13] [14]. Fourth, an important amount of analysis in MLG indicates the analysis via networks of various types of institutions, policy groups, advocacy groups as well as intergovernmental relations [15] [16]. The approach is regarded as institutionalist since it is aiming to understand the institution building aspect of europeanisation [17].

The major approaches to europeanisation studies can be grouped under two main theoretical frameworks. First, rationalist institutionalism is based on the logic of 'resource dependency' (p. 6, [4]). The Liberal Intergovernmentalist approach is concerned with the actors at the national level with an interest to rationally increase their autonomy [18]. Similarly in neofunctionalism, the supranational environment provides domestic actors with new resources that enable them to overcome their reliance on the central government [19]. In Multi-Level Governance, resource sharing is one way of motivation for actors to form policy networks as well as the issue of resource sharing is an important matter in the relations among actors at different levels of governance hierarchy [15]. The rationalist institutional theories in europeanisation offer two mechanisms in the mediation of resource dependency parameter. First, multiple veto points are decisive on the preference formation of domestic politics through arranging power sharing among actors.

The more power is disputed across the political system and the more the actors have a say in political decision-making, the more difficult it is to foster the domestic consensus or "winning coalition" necessary to introduce changes in response to Europeanisation pressures. (p. 7, [4])

Second, formal institutions enable actors with resources to elevate them for seizing further opportunities, enough to change the status quo [20].

On the contrary to rationalist variant, sociological institutionalism is based on the 'logic of appropriateness' [21]. Sociological institutionalism is also considered as the constructivist approach to europeanisation. Accordingly, in the formation of actor preferences, ideational resources, shared beliefs and ideas play central role [4]. Institutions are important not only because of the material resource dependency but also their norms, values and identities matter in the political context [22]. On top of all, the social context is formed in relation to interaction among institutional actors as social actors. Accordingly, sociological institutionalism is (1) *structuralist* which refer to the arising similarities among institutions that interact in shared environments; and (2) *agency-centred* which refers to the identity formation of actors via social learning and social interaction. Consequently, the mediating factors in sociological institutionalism are first, *change agents* or *norm entrepreneurs* who engage in the persuasion of domestic actors in the process which is called *social learning*. Second, *informal institutions* help shaping collective responses to europeanisation issues (pp. 8-9, [4]). Consequently, domestic change is possible at three levels, namely *absorption*, *accommodation*, and *transformation*. Absorption refers to the incorporation of European norms without changing the existing policy structure (low change). Accommodation refers to the adding up of new policy processes without essentially changing the existing policy structures (medium change). Transformation refers to replacement of existing policy structures by new ones (high change) [20]. In a similar way of analysis, Knill and Lehmkuhl (pp. 1-2, [23]) indicates that European institutions affect domestic policy structures towards transformation (*positive integration*); they constrain domestic preferences by restricting legislation (*negative integration*); or they transform the domestic policy environment dramatically (*framing integration*).

Constructivism allows the theoretical insight in europeanisation to focus on ideational actor preferences. Nevertheless it is possible to observe that it can embrace the MLG approach to europeanisation. The constructivist turn in MLG suggests that (1) it is not only the resource allocation in policy networks that is organised among the actors, but also the knowledge basis where interaction and communication is central in between actors. In the MLG analyses:

accumulation of knowledge, collective learning, and the exchange of ideas and concepts are significant. (p. 704, [24])

(2) An important contribution of MLG analysis is that it informs the process of decision-making is dispersed among actors on an horizontal basis thus the hierarchical governance structure no longer exist. In the confines of the MLG approach (3):

The organizing principle of political relations within the European system is based on consociation, which helps actors to manage heterogeneity within political communities. (p. 704, [24])

Different levels of governance interaction and dispersion of hierarchical centres of authority supports the assumption that consociation is an important feature for the organization of this specific type of governance. (4) The MLG makes it possible to incorporate the processes of *interaction* in addition to the two most important processes of europeanisation, favored by the LI and neofunctionalism, the processes of down-loading and up-loading of the European policies onto and from domestic polities (p. 13, [25]).

Empirically analysing the European system, it is also possible to locate the multi-level policy interaction that proceeds in a way predicted by constructivists. The process does not occur only in a one-way hierarchical manner (top-down), it also occurs bottom-up and mediated through social-institutional interactions along the way. The way member states influence European policy making system operates in four major steps or *points of access* (pp. 29-35, [26]). First, policies are designed at the European level through the deliberation of member state experts. Second, policies are negotiated among experts, ministers and governmental piers. Third, policies need to be legitimated to take effect in domestic polities, legally and practically. Fourth, policies are implemented with the addition of domestic flavour where the adaptation is influenced by the existing styles, beliefs, and understandings of the local policy makers.

By and large the EU system is characterised more by softer techniques than by the exercise of hard power, reflection of the character of the EU as a sophisticated form of cooperative multilateral and negotiated governance. (p. 36, [26])

Accordingly, we can bring up three types of governance mechanisms for europeanisation of domestic policies. First, governance by negotiation asserts that certain policies are negotiated among member states as well as supranational institutions. Second, governance by hierarchy refers to the enforcing of policies by the supranational institution, most of the times to limit the negotiating capacities of non-aligned members. Third, facilitated coordination refers to the lack of supranational initiatives apart from coordinating activities thus resulting in an atmosphere of horizontal consultation among members [27]. In constructivist variant of the MLG approach it is possible to reach to a single logic of europeanisation that can be applied over various policy areas [23] rather than the analysis of different policy areas with different logics.

2. Europeanisation and Enlargement

It is not much contested that Europeanisation is relevant not only for the member states of the European Union but also for the others who aspire to be part of it. The major difference in considering the europeanisation of member states of the EU and the candidates is the way European policies are transmitted into the domestic structures, whence in the second group it is more of a hierarchical process of intergovernmental bargaining. Thus, it is very relevant to talk about the europeanisation of enlargement. On the contrary, there has not been much inquiry towards the europeanisation of enlargement [28].

The 'Europeanisation effect' was therefore very strong externally – at least until the accession of most of those states became, for them, internal. (p. 340, 27)

Accordingly, the transformation of a country from intergovernmental negotiator towards a candidate negotiator hence into the europeanisation process is quite important.

The decision to accept more members into a supranational organisation, as a question of europeanisation, directs us towards the underlying logics. Analysing the recent enlargement towards Central and Eastern Europe (CEE), Schimmelfennig asserts that the driving logic for enlargement was far from material utilitarian position for optimising costs and realist presumptions of power politics [29]. On the contrary the europeanisation of new members occur according to a logic that assumes the appropriateness of prospective actors in relation to the supranational institutions in addition to material

bargaining. In other words, the ideas, belief structure and values of the domestic institution should comply with the supranational organisation prior to the process of europeanisation starts.

In a more in-depth analysis, Schimmelgennig and Sedelmeier (pp. 663-668, [30]) came up with a model of three governance structures. First, according to the *external incentives model*, intergovernmental bargaining appears in relation to the relative power of governments and their assessments of utility maximization. Second, according to the *social learning model* the logic of appropriateness rules the process which refers to the identities, values, norms of actors. "Correspondingly arguing about the legitimacy of rules and the appropriateness of behaviour (rather than bargaining about conditions and rewards), persuasion (rather than coercion), and 'complex' learning (rather than behavioural adaptation) characterizes the process of rule transfer and rule adaptation." (p. 667, [30]) Third, the lesson drawing model is based on the assumption that a rule from an outside source is implemented if it is expected to solve a domestic problem in an effective way. Thus, without challenging the existing theories of europeanisation, it is possible to draw outcomes for the enlargement process.

The most important challenge for the europeanisation of the enlargement process is the complexity of the process for the candidate state where the accession involve transformation of the whole system of policy making as well as identities and rule structures and the governance culture. As a result, the supranational organisation is in a position to charge candidate countries for the process of transformation of their current governance models.

[T]he EU accession process is pushing the applicant countries towards greater convergence with particular institutional models within the existing EU. (p. 1014, [31])

The process of europeanisation involves all structures of the government and state institutions. The whole system of governance and policy making mechanisms are expected to converge into the European model. At the basis of this process lies the *acquis communautaire* which is non-negotiable (p. 1015-1017, [31]). As a result, europeanisation models that assume the involvement of domestic institutions in a process of negotiation is not realistic. Nevertheless, the models based on the logic of goodness of fit can be favoured in the explanation of this process. Accordingly, the more a candidate state fits into the European model of policy making the less the transformation is expected to happen. However, it is not possible for the candidate state to resist change or influence the degree of fit.

Following on the analysis of Grabbe (pp. 1019-1024, [31]), five mechanisms of europeanisation can be presented concerning the enlargement process. (1) *Gate-keeping* function of the EU gives leverage to the supranational organisation during the negotiation process. It is the sole decision of the EU to elevate a candidate status in the stages of accession. Based on the process of accession during the CEE enlargement the europeanisation of enlargement follows certain steps. First, the countries were given privileged trade access and additional aid as foreign policy tools. Second, an enhanced form of association agreement was signed with the countries and expected to be implemented. Third, negotiations for accession was opened after the conditionalities for democracy and human rights were fulfilled. Fourth, 31 chapters on policy areas and governance were opened and closed depending on the progress of the countries. Fifth, accession treaty was signed to conclude the negotiation process. Finally, the treaty was ratified by the legislatures of each country. (2) *Benchmarking and monitoring* is based on the regular reporting on the country progress by the European Commission. It is an important tool to guide the processes at every stage in order to understand the progress of the europeanisation process. (3) *Models* provide the provision of legislative and institutional templates to be implemented by candidate countries. Models work as roadmaps for the necessary changes to be implemented at the domestic level. Models are based on the *acquis communautaire* but also involve advisory points on further development. (4) *Money* was an instrument as an incentive as well as a market mechanism. (5) *Advice and twinning* practices that indicate the process of learning by state employees and agents. As a result of the europeanisation in CEE, Grabbe (pp. 1028-1029, [31]) draws attention to three major impacts on domestic politics. First, the balance of power in the political system is affected. It is said that the legislature is marginalized since the power and resources are concentrated in the executive who has key responsibilities in the process of negotiations. Second, a *core executive* emerges as a result of the process further creating imbalances even inside the executive branch. Third, sub-state

elites are excluded from the decisions taken on behalf of them during the creation of sub-state reform. An important concern arising in these circles is the increasing importance of efficiency at the expense of legitimacy. Furthermore, the process is used to legitimise institutional frameworks and preferences over governance models.

3. Impact of Europeanisation in South East Europe

In South East Europe (SEE) the europeanisation can be considered in a similar way it had occurred in CEE. The path-dependency of europeanisation is relevant for SEE countries in their relation to EU institutions. In addition to the mechanisms, mentioned above we should take the special status of SEE into consideration as a conflict laden area. First, *coercion* is exercised via control and conditionality by the EU. Second, *mimetism* is practiced via contagion which refers to the “dissemination of experience” and consent (pp. 89-90, [32]). Apart from the normal channels of EU international relations with the countries of SEE, control means the use of positive and negative sanctions. In relation to this, conditionality is the type of coercive activity where the benefits are distributed on the basis of compliance to specific conditions, akin to country-specificity. Contagion refers to the process of learning and consent refers to the inclusion of the domestic population’s participation in contact with the supranational context. However, the specific conditions of SEE in its relations with Europe are quite crucial in any kind of analysis of europeanisation. The existing political, economic, social and institutional conditions inhibit normal relations with supranational institutions. Thus the involvement of the European institutions carries also an intention for correcting mechanisms for the existing conditions [32]. The course of the relations between SEE countries and the EU has been shaped under the influence of crisis situations. Following this argument the following assumption can be straightforward. The relations between EU and the SEE are shaped with a concern for credibility in the region for the European institutions [33]. The lessons drawn prior to the Kosovo crisis indicated that the prospect for membership was the only way of containing the region with peaceful and democratically accountable countries. Europeanisation, thus can be interpreted as a problem-solving apparatus for the whole region as a major political project [32].

Similar to the process of europeanisation in CEE, first, certain countries in SEE¹ are offered trade privileges, financial aid and contractual benefits as a first step of rapprochement. The Royamaunt process that started in 1996 enabled this set of relations in a process of regional approach. However, the effectiveness of the approach and the slowness of the process was proved as of the Kosovo crisis, after which the europeanisation was speeded up. Second, a Stabilization and Association Agreement was offered to each country by the EU to define a common European approach on the question of regional instability. SAAs provided the involvement of a membership perspective for the countries of SEE although at an undefined timetable [33]. The Stabilization and Association Process (SAP) is expected to take a long time for SEE considering the existing conditions of the countries. Moreover, the continuing material and ideational support is expected to continue for the whole process to be successful [32].

An important feature of the europeanisation process that we can also observe in SEE is the adoption of variable geometry, although there is a regional policy of the EU on this matter. The variable geometry concept assumes that different negotiating states at different levels of their development as well as their progress during negotiations might move into coming stages independently of each other [34]. In SEE this is more relevant since problems of political conditionality – human rights, democracy – and economic conditionality are major challenges in the region. According to the variable geometry approach, embedded in Stabilization and Association Process (SAP), as each country progresses towards achieving pre-set goals of the European institutions, it is rewarded and encouraged for further steps. SAP aims at signing of SAA among the countries in the region hence creating a free trade area with the EU. Through the SAAs, the support for democratic consolidation, rule of law, economic

¹ For the purposes of accuracy here, the countries of western Balkans, namely Albania, Croatia, Bosnia Herzegovina, Federal Republic of Yugoslavia (later Serbia and Montenegro), and FYR Macedonia are considered SEE countries. Bulgaria and Romania were late followers of CEE in the process of europeanisation as candidate countries; geographically in SEE, Greece is a EU country.

development and regional cooperation are aimed. The SAP involves material aid from the European Community in terms of aid under the framework of CARDS (Community Assistance for Reconstruction, Development and Stabilization) program [35]. The priorities of the CARDS program are concentrated on institutional building for the countries in SEE towards further integration. The previous arrangement of financial aid for the region (until 2000 – including Albania, FYR Macedonia, Croatia) was structured along the PHARE program together with other CEE countries in which the aim was to provide aid in terms of developing infrastructure [36] in addition to the administrative capacity building [37].

The SAP and relevant SAAs can be considered as relevant policy tools for the SEE countries towards European integration. As in the case of integration process, each country is considered in relations to its own conditions at the same time a regional approach is developed.

The conclusion of a SAA is viewed as a result of the efforts of the country concerned to fulfil the relevant political economic criteria and as a recognition of the credibility of the country's aspiration to become a candidate for full EU membership. At the same time the punitive aspects of conditionality continue to play an important role and are an integral part of each SAA. (p. 235, [38])

The SAA and its basis on TEU and the Copenhagen criteria essentially tend to direct the efforts of intergovernmental relations in the context of European integration. As most important conditionality requirements are based on the issues of human rights, rule of law and democratisation, European instruments during the SAP are operationalised through sticks and carrots approach. Leverage on the CARDS program and benchmarking activities of the Commission are major European instruments in this sense [38].

4. FYR Macedonia and the Framework Agreement

On January 2001, a paramilitary group called National Liberation Army (NLA) claimed the responsibility for an attack to a police station to the north of Macedonia close to the border with Kosovo. It was believed that ammunition and militia had been freely smuggled via the region due to the lack of an actual border between Kosovo and Macedonia, as well as the border police. The NLA was thought to be an offshoot of the Kosovo Liberation Army, which was then passive. At first, the group was underestimated by the Macedonian authorities but the insurgency gained momentum, spreading into Tetovo in March. The group gained support from the local Albanian population that complicated the matters. The military operations by the state were not successful. Due to large Albanian support for the fighters the Macedonian authorities started to arrest a great number of Albanians. The tension between two ethnic groups widened immeasurably. As a result of the guerrilla warfare there were around 70,000 internally displaced persons from both sides. The NLA pushed for the rights of Albanians which had been on the agenda of Albanian political parties for more than a decade. Contrary to Macedonian allegations, NLA denied a demand for secession. With the involvement of international negotiators a cease-fire was effective as of July 9th, 2001 followed by the negotiations for an Agreement to end the conflict – known as the Ohrid Agreement or the Framework Agreement [39]. The purpose of the Ohrid Agreement also known as the Framework Agreement was to ease the tension between ethnic Macedonian majority and the Albanian minority in Macedonia. Internationally it was a mechanism to contain the spill over of the Albanian extremism from bordering Kosovo into conflict-laden Macedonia. The document is produced under the auspices of the President Boris Trajkovski and with the help of international negotiators; James Pardew (USA) and François Leotard (EU) [40]. The involvement of international actors in the process of negotiations was crucial. The existence of incentives such as the Stabilisation and Association Agreement from the part of the EU for the signing of the agreement helped the conflicting sides to come into terms for Agreement. It is viable to argue that an incentive as big as European Union accession allowed the conflicting parties to give up zero-sum arguments of ethnic nature [41].

In brief terms, the Agreement aimed at satisfying Albanian demands as to secure their minority rights at the same time preserving the unitary character of the state. The Agreement pointed to the necessary changes in the constitution that implied the ownership of the state to belong to only one ethnic

community. So far, Albanians would be equal to ethnic Macedonians in the wording of the constitution. Second, it elevated the Albanian language to official state language status. Third, it empowered the Albanian minority with veto rights regarding certain policy areas: culture, use of language, education, personal documentation, the use of symbols, the laws on local finance, local elections, the boundaries of municipalities and the city of Skopje, the law on local self-government. Fourth, the Agreement addressed Albanian under-representation and attempted to devolve central authority through decentralization [42]. Most important asset of the agreement is the decentralisation process that involves increasing consociation of the local governance. The agreement addressed necessary changes at the local level for the state to become more European via civic ideals.

Ohrid stipulated a certain degree of devolution of power from the central state to municipalities. This had been a main Albanian demand, but as an important administrative reform in its own right, it is also considered a key element in the country's security framework and its EU prospects. (p. 4, [43])

The implementation of the law on local government, in this sense, is an instrument to bring the country closer to Europe as well as an instrument that brings the Albanian minority the opportunity for greater political autonomy where they are numerous.

The major issues of the Agreement were the status of the Albanian language and the composition of the security forces in Macedonia. Regarding the language issue, Macedonian state officials argued that acceptance of Albanian as official language beside Macedonian would result in the federalization of the state undermining its unitary character². According to Albanians though,

a significant portion of their demands overlapped with political reforms that the Macedonian state was supposed to be implementing as part of the EU Stabilisation and Association Process. (pp. 30, [40])

During the negotiations, the middle way was found: Any language that is spoken more than 20 percent of the population would be considered official language beside Macedonian which would be the only official language regarding international matters. This would mean automatically that Albanian would be elevated to the official language status without the recognition of it 'equal' to Macedonian that might symbolically harm the unitary character of the state.

Second, the Albanian side demanded the localization of the security forces. In other words, they asked for the local police authority to be placed under the jurisdiction of local authorities. The major reason for this demand lies in the violent clashes between the police forces and the Albanian population. Macedonians observed protests against the police force as

manifestations of a broader resistance to all state authority. (n6, pp. 149, [44])

According to the Macedonian side, localization of the security forces would threaten the unitary character of the state. Also another problem would be the creation of various security forces in different parts of the country depending on the ethnic composition. Practically, an all-Albanian police force in north and north-west parts would jeopardise the non-Albanian population according to the Macedonian side. The solution was founded however both reflecting the ethnic composition at the same time preserving the unitary character of the police force. Accordingly, the ethnic composition of the police would reflect the ethnic composition of the country. Thus, a certain amount of Albanians would be recruited by the police force. A second solution regarding the police authority was founded according to the previous practice during the Yugoslavia period: The local police department chief would be chosen by the municipal council from a list offered by the ministry of interior³ [40].

In general, the Agreement was considered successful thanks to the end of the armed insurgency and decreased amount of inter-ethnic violence. Nevertheless, a certain amount of violent activities were

² "Macedonia's sovereignty and territorial integrity, and the unitary character of the State are inviolable and must be preserved. There are no territorial solutions to ethnic issues." (Article 1.2, [45])

³ In order to ensure that police are aware of and responsive to the needs and interests of the local population, local heads of police will be selected by municipal councils from lists of candidates proposed by the Ministry of Interior, and will communicate regularly with the councils. The Ministry of Interior will retain the authority to remove local heads of police in accordance with the law. (Article 3.3, [45])

observed in the aftermath of the Agreement. The violent events in 2002 and 2003 especially in the regions where Albanians are numerous and in the capital, Skopje did not turn out to gain ethnic character. According to ICG (pp. 4-7, [44]) the main reason for continuing violence in the previous crisis areas was (1) the inability of the police force which had been restructured according to the Framework Agreement and, (2) the lack of confidence among the Albanian population for Macedonian state institutions, including the police.

The Agreement gained respect between the major two ethnic groups as the country moved towards the implementation of the decentralization process and the progress achieved in the international arena in connection with this.

The application Macedonia submitted for European Union membership on 22 March [2004] is of tremendous importance for stability. The prospect of EU integration gives politicians their main motivation for pursuing reform policies and helps guarantee peaceful coexistence of the main ethnic groups. (p. 1, [43])

As during the signing process, the intervention by the international community during the implementation process had been considered vital.

The whole process of implementing the Framework Agreement proved that external monitoring, support and occasionally intervention is crucial for the realization of the planned reforms because, without international mediation, the political parties in the Republic of Macedonia hardly find compromises on those vital issues. (p. 7, [46])

The case of the Framework Agreement clearly showed that the existence of international involvement was vital and necessary from the beginning. First, it is obvious that the major reason for the parties to start negotiations and then later to come into an agreement was due to the pressure exercised by the representatives Pardew (USA) and Leotard (EU). Specifically, the EU had the ability to force the parties towards agreement due to the country's existing situation as a SAP country. Second, the legitimacy of both sides were guaranteed in the eyes of the international community making them not only domestic negotiators but also piers in front of the international community responsible for their constituents. The negotiating process was a process of two-level diplomatic game [47] in one sense and a more complicated process to involve intergovernmental bargaining. Third, the pressure from European institutions was apparent not only in the form of coercion to force parties for a conflict settlement. However, as the course of the Agreement suggested the development of a policy initiative, namely the decentralization process, instrument of financial aid is pronounced. Hence the European involvement in the process was accomplished through a carrots and sticks approach. Fourth, the resulting progress towards stabilization was awarded by the supranational organization in the form of SAP initiatives as well as a membership candidacy perspective [48].

From a functionalist point of view, the existence of the europeanisation instruments filled the void of legitimacy for the unitary character of the country to stay intact. It is not possible to imagine FYR Macedonia to stand outside the framework of europeanisation with material as well as ideational support from European institutions. The process of institution building in line with European models of consociation and local governance was possible with the formation of an identity structure, which is European, multi-level, and internationally legitimate. The question of direct domestic legitimacy was surpassed via the mechanisms of international involvement in the process. The whole structure thus, becomes credible in the eyes of the domestic constituency due to the existence of the auspices of the international community. Additionally, elites consolidate their domestic positions of bargaining for consociation with the approval of the European framework. This indicates a dilemma between the necessity of institution building, and administrative capacity and the democratic capacity of the public administration. The same dilemma had been observed in the CEE integration and caused concerns [37].

5. Conclusion

The theoretical approaches to europeanisation are relevant to analyse the political phenomena in SEE although the states in the region suffer from continuous instability, lack of institutionalisation, democratic accountability, problems associated with post-communist transition, and developmental issues. All of the above mentioned structural problems are of interest to politics of transition, political economy, international relations and comparative politics. Europeanisation might help to create a framework to analyse the complexity involving theoretical insights from all these disciplinary approaches.

I tend to consider the case of Macedonia in terms of europeanisation analysis from the point of view of constructivist institutional approach to Multi-Level Governance. The reasons are: (1) The emergence of new and multi-level governance mechanisms with emphasis on European institutional compatibility and consociation, (2) the development of the institutional context after a decade of instability and economic deprivation with endemic problems of democratisation, development and good governance, (3) the importance of the Europeanisation framework with an identity setting function via providing legitimacy for reform and institution building. Developments occurred after the signing of the Framework Agreement indicated that the process has been more than an intergovernmental bargaining process but it incorporated a vision of European governance structures. The usage of membership perspective for the near neighbourhood of Europe was consolidated after its introduction in 1999 via the SAP.

References

- 1 Buller, J. and Gamble A. Conceptualising Europeanisation. *Public Policy and Administration* 2002, 17(2)
- 2 Vink, M. What is Europeanisation? And Other Questions on a New Research Agenda. *European Political Science* 2003, 3(1)
- 3 Buller J. The Europeanization of National Politics? *Government and Opposition* 2003, 38(4)
- 4 Boerzel, T. A. and Risse, T. When Europe Hits Home: Europeanization and Domestic Change. *European Integration Online Papers* 2000, 4(15) <http://eiop.or.at/eiop/texte/2000-015a.htm> (viewed on: 15/02/2006)
- 5 Risse, T., Cowles, M. G. and Caporaso, J. "Europeanization and Domestic Change: Introduction" in M. G. Cowles, J. Caporaso and T. Risse (eds.) *Transforming Europe: Europeanization and Domestic Change* 2001, NY: Cornell University Press.
- 6 Olsen, J. P. The Many Faces of Europeanization. *Journal of Common Market Studies* 2002, 40(5)
- 7 Wallace, H. Europeanisation and Globalisation: Complementary or Contradictory Trends? *New Political Economy* 2000, 5(3)
- 8 Ladrech, R. Europeanization of Domestic Politics and Institutions: The Case of France. *Journal of Common Market Studies* 1994, 32(1)
- 9 Boerzel, T. Towards Convergence in Europe? Institutional Adaptation to Europeanization in Germany and Spain. *Journal of Common Market Studies* 1999, 39(4)
- 10 Radaelli, C. "The Europeanization of Public Policy" in K. Featherstone and C. Radaelli (eds.) *The Politics of Europeanization* 2003, Oxford: Oxford University Press.
- 11 Stoker, G. Governance as Theory: Five Propositions. *International Social Science Journal* 1998, 50(155)
- 12 Smith, A. Studying Multi-Level Governance. Examples From French Translations of the Structural Funds. *Public Administration* 1997, 75, (Winter)
- 13 Sharpe, L. J. The Growth and Decentralization of the Modern Democratic State. *European Journal of Political Research* 1988, (16)
- 14 Smith, B. C. *Decentralization: The Territorial Dimension of the State* 1986, London: George Allen and Unwin.
- 15 Rhodes, R. A. W. *Understanding Governance: Policy Networks, Governance, Reflexivity and Accountability* 1997, UK: Open University Press.

- 16 Bache I. and M. Flinders "Themes and Issues in Multi-Level Governance" in I. Bache and M. Flinders (eds.) *Multi-Level Governance* 2004, NY: Oxford University Press.
- 17 Jordan, A. *The European Union: An Evolving System of Multi-Level Governance ... or Government? Policy and Politics* 2001, 29 (2)
- 18 Moravcsik, A. Preferences and Power in the European Community: A Liberal Intergovernmentalist Approach. *Journal of Common Market Studies* 1993, 31(4)
- 19 Sandholtz, W. Membership Matters: Limits of the Functional Approach to European Institutions. *Journal of Common Market Studies* 1996, 34(3).
- 20 Boerzel, T. and Risse, T. "Conceptualizing the Domestic Impact of Europe" in K. Featherstone and C. Radaelli (eds.) *The Politics of Europeanization* 2003, Oxford: Oxford University Press.
- 21 March, J. and Olsen, J. *The Institutional Dynamics of International Political Orders*. International Organization 1998, 52(4)
- 22 Bulmer, S. and Lequesne, C. "The European Union and its Member States: An Overview" in S. Bulmer and C. Lequesne (eds.) *The Member States of the European Union* 2005, NY: Oxford University Press
- 23 Knill, C. and Lehmkuhl, D. How Europe Matters. Different Mechanisms of Europeanization. *European Integration Online Papers* 1999, 3(7) <http://eiop.or.at/eiop/texte/1998-007a.htm> (viewed on: 15/02/2006)
- 24 Knodt, M. International Embeddedness of European Multi-Level Governance. *Journal of European Public Policy* 2004, 2(4)
- 25 Howell, K. Developing Conceptualizations of Europeanization and European Integration: Mixing Methodologies. ESRC Seminar Series 1, UACES Study Group 2 presented on 29/11/2002 in University of Sheffield, UK.
- 26 Wallace, H. Exercising Power and Influence in the European Union: The Roles of Member States. in S. Bulmer and C. Lequesne (eds.) *The Member States of the European Union* 2005, NY: Oxford University Press
- 27 Bulmer, S. J. and Radaelli, C. M. in S. Bulmer and C. Lequesne (eds.) *The Member States of the European Union* 2005, NY: Oxford University Press
- 28 Schimmelfennig, F. and Sedelmeier, U. Theorizing EU Enlargement: Research Focus, Hypotheses, and the State of Research. *Journal of European Public Policy* 2002, 9(4)
- 29 Schimmelfennig, F. The Community Trap: Liberal Norms, Rhetorical Action, and the Eastern Enlargement of the European Union. *International Organization* 2001, 55(1)
- 30 Schimmelfennig, F. and Sedelmeier, U. Governance by Conditionality: EU Rule Transfer to the Candidate Countries of Central and Eastern Europe. *Journal of European Public Policy* 2004, 11(4)
- 31 Grabbe, H. How does Europeanization Affect CEE Governance? Conditionality, Diffusion and Diversity. *Journal of European Public Policy* 2001, 8(6)
- 32 Demetropoulou, L. Europe and the Balkans: Membership Aspiration, EU Involvement and Europeanization Capacity in South East Europe. *South East European Politics* 2002, 3(2-3)
- 33 Lykke, F. and Murphy, A. Turbo-Charged Negotiations: The EU and the Stability Pact for South Eastern Europe. *Journal of European Public Policy* 2000, 7(5)
- 34 Goldsmith, M. Variable Geometry, Multilevel Governance: European Integration and Subnational Governance in the New Millennium. *The Politics of Europeanization* 2003, June.
- 35 Hoffman, J. Integrating Albania: the Role of the European Union in the Democratization Process. *Albanian Journal of Politics* 2005, 1(1)
- 36 Alkan, N. "Borders in Europe" in N. Alkan (ed.) *European Integration and South Eastern Europe* 2002, Bonn: Center for European Integration Studies.
- 37 Dimitrova, A. L. Enlargement Governance and Institution Building in Central and Eastern Europe: The Case of European Union's Administrative Capacity Requirement. Paper presented on 6-8/09/2001 in ECPR General Conference at the University of Kent, Canterbury
- 38 Pippan, C. The Rocky Road to Europe: The EU's Stabilisation and Association Process for the Western Balkans and the Principle of Conditionality. *European Foreign Affairs Review* 2004, 9
- 39 Roudometof, V. *Collective Memory, National Identity and Ethnic Conflict: Greece, Bulgaria, and the Macedonian Question*, 2002, CT, USA: Praeger

- 40 Popetrevski, V. and Latifi, V. "The Ohrid Framework Agreement Negotiations", The 2001 Conflict in FYROM – Reflections, 2004, UK: Conflict Studies Research Center (June)
- 41 Piana, C. The EU's Decision-Making Process in the Common Foreign and Security Policy: The Case of the Former Yugoslav Republic of Macedonia. *European Foreign Affairs Review* 2002, 7
- 42 Dimitrova, K. R. Municipal Decisions on the Border of Collapse: Macedonian Decentralization and the challenges of Post-Ohrid Democracy. *South East European Politics* 2004, 5(2-3)
- 43 ICG. Macedonia: Make or Break. *Europe Briefing* 2004, August 3.
- 44 ICG. Macedonia: No Room for Complacency. *Europe Report* #149 2003, October 23.
- 45 Framework Agreement 13.08.2001
<http://www.president.gov.mk/prilozi/dokumenti/180/FRAMEWORK%20AGREEMENT.pdf>
 (viewed: 10/06/2005).
- 46 Brunnbauer, U. The Implementation of the Ohrid Agreement: Ethnic Macedonian Resentments. *Journal of Ethnopolitics and Minority Issues in Europe* 2002, 1
- 47 Putnam, R. D. Diplomacy and Domestic Politics: The Logic of Two-Level Games. *International Organization* 1988, 42(3)
- 48 Council of the European Union. Presidency Conclusions: 30/01/2006, Brussels.
http://ue.eu.int/ueDocs/cms_Data/docs/pressData/en/ec/87642.pdf (viewed on 18/05/2006).

Effects of Europeanisation on Social Policies of Slovenia and Macedonia: Convergence vs. Disparity?

Maja Gerovska Mitev,

SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, mgerovska@seerc.info

The European Union has become entrenched as an important supranational player in social policy creation. Although it initially tread into this area cautiously and sporadically and is dependent on its economic policy, it now seems that the language used by Open Method of Coordination (OMC), Joint Inclusion Memorandum (JIM), National Action Plan for Employment (NAPE) and European Employment Strategy (EES), shake and lead the way in social policy creation in most European countries. But what do these mechanisms achieve? What do “hard” and “soft” mechanisms bring to national social policies? Can these mechanisms compete with the others that have been introduced by other internationals? In an effort to address these questions, this article intends to explore the applicative aspect of Europeanisation in social policies. In so doing, this article begins by outlining the possible effects of Europeanisation in the institutional, regulative and cognitive spheres, which may possibly lead to convergence, divergence and disparity as outcomes. This is followed by a brief exploration of EU influence on the social policies of the newer member states, which shows that the social policies of EU 10 seem to be more “Europeanised” than those of EU 15, primarily due to the need of the former to adapt and alter their previously ineffective welfare models. However, the main empirical evidence in this article is drawn from an analysis made of the social policies of Slovenia and Macedonia, where the influence of external factors and the decisiveness of the internal players indicate these to be the most prominent agents for welfare reform. The external factors are not only analysed through EU impact, but also through the actions of the Breton Woods institutions (World Bank and IMF), which in some cases have managed to substitute and outplay the EU in the field of social policy prescription. The concluding part of this article intends to demonstrate the various effects of Europeanisation in these two countries, thereby indicating the prevalence of convergence or disparity in the enlarged European Union. The abstract should be around 150-200 words and definitely fit in the first page of the paper. This document is a sample document that gives the authors of SRC 2006 a general overview of how their camera-ready paper should look like. This paper will be included in the Proceedings of the conference published by SEERC. The document is based on the styles created in the template available from the site of the conference.

Keywords

Europeanization, EU social policy, convergence, divergence, Slovenia, Macedonia.

1. Possible Effects of EU Social Policy Dissemination

European social policy goals and standards have been conveyed through the dissemination of various instruments, which are directed at both member and candidate countries. These instruments are comprised of a limited number of ‘hard’ mechanisms (social acquis), as there has been general opposition for more of these social policy instruments and numerous ‘soft’ mechanisms (OMC, NAPE, NAPs etc.). The prevalence and unobtrusiveness of the “soft” social policy mechanisms have succeeded in achieving significant results. The outcomes of these mechanisms will be assessed through the innovations and modifications seen on the institutional, legislative and cognitive level,

which may therefore contribute to the evidence of convergence and/or divergence and disparity among the EU and member or candidate countries.

1.1 Institutional and Regulative Europeanisation

Europeanisation can be observed both as a process and as an outcome in national social policies. According to Ioakimides, Europeanisation can influence the redefining of national political systems on four levels: institutional, regulative, functional and territorial [1]. Concerning the outcomes in national social policies, it can be argued that the most visible changes are those on the institutional and regulative level. Institutional changes are mainly visible through modifications made to the bodies who administer national social policy (renaming and/or creating new departments within Ministries of Social Affairs, changes in the Employment Bureaus/Agencies, etc), but they can also be observed through the creation of completely new institutions (i.e. Public Guarantee Fund, Office for Equal Opportunity, etc.), which complement and improve the administrative capacity of national social protection systems.

On the other hand, regulative Europeanisation has a more distinctly visible outcome. In some ways, it is a ‘must’ in the process of negotiating social *acquis* chapters, and therefore many of the ‘hard’ legislative criteria has become embedded in national social policy legislation. This mainly involves the need to incorporate numerous labour law directives, health and safety at work directives, as well as equal opportunity directives. Other, more ‘formal’ domains of social policy are regulated through ‘soft’ mechanisms, and it is also through input from these channels that has resulted in the creation of national action plans for employment, joint inclusion memorandum, etc.

A summary of possible institutional and regulative effects that are supported through various EU mechanisms is presented in Table 1. Although the institutional and regulative effects can differ among countries depending on their capacity and need to implement these changes, it can be argued that the EU favors an active social policy, accompanied by more direct public authority.

Instruments	Effects	
	Institutional Europeanisation	Regulative
Open Method of Coordination	Activation of local and regional levels in the coordination of national social policies	National Action Plan for employment; National Action Plan for social inclusion
Structural Funds (ESF)	Modernisation of systems for education, training and employment	Adoption and streamlining of European Employment Strategy priorities
Social Acquis	Creation of new institutions for administration of social policy; Creation of separate sectors for coordination of European affairs within Ministries for Social Policy.	Adoption of European directives in national legislation concerning: Free movement of workers and coordination of social security; Employment and social policy; Education, Training and Youth
Pre-accession programmes (PHARE, CARDS)	Transformation of local agencies for employment; Modernisation of local social services	Creation of new criteria for social assistance payments; Initial creation of national action plans for employment

Table 1 Institutional and regulative Europeanisation of national social policies

1.2 Cognitive Europeanisation

Besides its direct effects, the EU influence in national social policies can be observed through indirect effects. According to Radaelli, indirect effects can be analysed through the process of cognitive Europeanisation, a process that affects both perception and attitude towards social problems, as well as the manner in which these are managed [2]. Measuring cognitive Europeanisation is much more difficult than measuring institutional or normative Europeanisation, but as Palier and Guillen note, the

cognitive aspect is an important one when analysing changes in public debate, logic and discourse, because it facilitates the way for current and future political and institutional changes [3]. Similarly, Lendvai associates this process in social policy with the appearance of new terminology, new meanings, new agendas and discourse that become visible during the EU accession process [4].

However, if one attempts to analyse the effects of cognitive Europeanisation, it may be contended that those changes would be more visible in EU (member and candidate) countries where social policy does not have a long tradition (i.e. South-Eastern European countries, such as Portugal, Greece, Spain and Italy), as well as in those countries where social policies were under the auspices of autocratic regimes (such as the ex-socialist countries of Central and Eastern Europe). In this group of countries, the strife to promote the European Social Model and its goal of solidarity and efficiency can be seen more clearly because of the different values of the immanent social protection systems in the past.

1.3 Convergence

Convergence as an effect in social policies can be observed as a correlation of national social policies towards the EU preferred goals or that of the European Social Model. Within the given literature, the thesis of convergence among social policies finds a similar number of supporters and opponents. Neo-functional representatives within the Union argue that the creation of the Single European Market contributes to a convergence of social protection levels within the member states. According to Falkner [5], the economic integration will inevitably contribute to functional ‘spillover,’ especially in social policies. Bertozzi and Bonolli, on the other hand, indicate many factors that condition the convergence of national social policies, such as political and organisational culture/tradition, the empowerment of different players, as well as the learning process [6].

Measuring convergence in practice proves extremely difficult because, as was rightly pointed out by Börzel and Risse, what appears as a convergence on a macro level may show a significant degree of divergence on a micro level [7]. Therefore, in order to specify convergence in a more measurable manner, this article will analyse: a) convergence of policies; b) convergence of instruments and c) convergence of effects. Table No. 2 indicates various dimensions and degrees for measuring convergence that will be used in the selected case studies.

Dimensions	Degree of convergence		
	High	Middle	Low
Convergence of policies	Applicative and normative compatibility with policy goals for: employment, social inclusion and pension	Presence of normative but an absence of applicative compatibility with policy goals for: employment, social inclusion and pension	Declarative efforts with no real evidence of applicative and normative compatibility with policy goals for: employment, social inclusion and pension
Convergence of instruments	Use of same mechanisms for realisation of social policies	Use of similar mechanisms for realisation of social policies	Use of alternative mechanisms for realisation of social policies
Convergence of effects	Compatibility of social transfer expenditures; employment/unemployment rate; poverty rate	Similarity of social transfer expenditures; employment/unemployment rate; poverty rate	Disparity of social transfer expenditures; employment/unemployment rate; poverty rate

Table 2 Dimension and degree of convergence in national social policies

On the basis of this approach, an analysis of the social policies of Slovenia and Macedonia will be made in order to determine the level of their convergence with the EU social policy standards.

1.4 Divergence

Divergence supporters highlight many arguments which defend the prevalence of divergence among European national social policies. Firstly, they emphasize the existence of different traditions and diverse political preferences in the creation of social policy in Europe, which presents an important obstacle on the path to convergence. Furthermore, additional factors that contribute to the divergence scenario are the limited number of EU social policy domains, their subsidiary status and the limited capacity of supranational institutions to dictate the course of national social policy. Finally, as Kleinman notes, “increased top down pressures for convergence for the sake of the European project may provoke a centrifugal reaction, leading to demands for greater divergence and diversity in the future” [8].

On the basis of these arguments, it may be concluded that divergence is an obvious effect, and the dominance of national tradition over international influence in social policy is a prevalent feature. Yet, the supporters of divergence do not present any contra evidence on the ‘spillover’ arguments, which, accompanied with the effects of the economic monetary union and other ‘unintended’ and collateral consequences of European integration may describe the convergence scenario as a possible one.

Measuring divergence can be achieved by comparing the difference in the degree and level of social expenditure and public intervention, as well as the diversity of goals and its differing effects on social policies in EU (member and candidate) countries.

1.5 Disparity

Disparity as an effect is similar to convergence, but the emphasis here is not put on the non-compatibility of national social policies with those on the EU level, but rather on the lack of similarity and prevalence of totally different social policy solutions. As such, disparity is indicated when basic values and manners of social policy creation and administration are inconsistent with the European norms. The prevalence of disparity depends on various rooted values, such as tradition, national priorities and dominant political forces that contribute to the creation and provisions of social policies.

An analysis of disparity may be observed through the existence of various socio-economic conditions in the EU (member and candidate) countries, and hence, disparity can be analysed as both an effect, as well as a factor that influences the attainment of convergence in the social goals of the Union.

2. Influence of European Social Policy in the Newer EU Member States (EU 10)

One of the main challenges that the EU has been confronted with during its existence has been its enlargement to include the countries of Central and Eastern Europe. In regards to the social policy aspect, the EU 15 had been skeptical about these enlargement prospects for fear of: a) social dumping; b) social tourism and c) the take up of available jobs. However, these worries never materialised, as the EU 15 had undertaken certain preventive measures to protect their welfare systems from the ‘Easterners’.

From the system perspective, however, an important attribute that these countries brought with them was a different social policy model. While all the new EU member states do not have one homogeneous model, there are certain characteristics which are shared by all of them and which are most likely to change due to the need for the social preferences of the Union. Some of these shared characteristics include:

- A bureaucratic welfare state, which provides social services in a centralized and non-transparent manner.
- Lack of experience in the private sector as a provider of social services.
- Ineffective social partnerships and social dialogue in the creation of national social strategies.

- Emphasized egalitarian expectations of the citizens in provisions of generous social benefits by the state.
- Ineffective and non-targeted social benefits.
- No correlation between social and economic policy.
- Immeasurable goals and an emphasis on declarative statements for social programmes.

Other social, economic and demographic characteristics of the new EU member states also indicated certain differences with those of EU 15.

Country	2001	2003
EU 15	7.4	8.1
Cyprys	4.4	4.4
Czech R.	8.0	7.8
Estonia	11.8	10.1
Hungary	5.6	5.8
Latvia	12.9	10.5
Lithuania	16.1	12.7
Poland	18.5	19.2
Slovenia	5.8	6.5
Slovakia	19.5	17.1
Malta	6.7	8.2

Table 3 Unemployment rate (%) in the new EU member states (EU 10) Source: European commission, 2004

Country	Before social transfer	After social transfer
EU 15 (2001)	39	15
Cyprus (1997)	24	16
Czech R. (2001)	36	8
Estonia (2002)	42	18
Hungary (2001)	44	10
Latvia (2002)	43	16
Lithuania (2001)	41	17
Poland (2001)	48	15
Slovenia (2002)	37	11
Slovakia (2002)	43	21
Malta (2000)	30	15

Table 4 Share of Poverty Rate Source: Eurostat, Laeken indicators 2004

From the figures given, it can be noted that there are principle differences in the area of unemployment, where only four countries (Cyprus, Czech Republic, Hungary and Slovenia) are lower or similar to the EU 15 rate of unemployment. This indicates the need to give priority to new approaches and solutions for the creation of employment policies in the countries of Central and Eastern Europe. In the remaining domains, the new member states do not lag behind significantly. In terms of poverty rates, those much higher than the EU 15 are only Slovakia, Estonia and Lithuania. The period of transition has obviously had a negative impact on social expenditure levels, where it can be comparatively seen that all new member states have had lower social expenditures than those of the EU 15.

During the period of negotiations with the European Union, the new member states were primarily faced with the necessity of adapting their social legislation with that of the European Union. The transposition of the social *acquis* has not created great problems, but in the area of the free movement of labour, where certain restrictions in the older member states apply, transitional periods have been formed (3+2) for utilizing the full capacity of labour markets in the European Union by the workers of the new member states. What has made a greater impact in terms of change and adaptation in national social policies has been the ‘soft mechanisms’ of the Union, which in the domains of employment and social inclusion have proven to be most effective.

Country	2002
EU 15	28.0
Cyprus	16.5
Czech Republic	19.9
Estonia (2001)	14.3
Hungary	20.9
Latvia (2000)	17.8
Lithuania (2000)	15.8
Poland (2001)	22.1
Slovenia	25.4
Slovakia	19.2
Malta	17.7

Table 5 Social expenditures as % of GDP Source: European Commission and Eurostat

Within the employment policy of the (then) candidate countries, initial steps were taken through the creation of the Joint Assessment of the Employment and Labour Market. These reports recognised the main challenges faced by the labour markets, which then contributed towards the creation of National Action Plans for Employment, where more practical steps were planned for the undertaking of necessary activities. According to the European Commission Report, national action plans have improved national policies in certain areas such as: incentives for making the work pay off; reforms in public agencies for employment; design and evaluation of strategies for life-long learning and facilitation and development of businesses [9].

Social inclusion policies have also been a part of the Europeanisation trend. Cooperation in this area was initiated through the creation of the Joint Memorandum for Social Inclusion (JIMs), which has had a positive impact on the adoption of quantitative indicators to determine poverty and social exclusion. For the first time, these countries embraced comparative and measurable standards, which in the future will enable the assessment of the effectiveness of their social inclusion policies. Further itemization of the JIMs came in the form of National Action Plans for Social Inclusion (NAPs), which were based on the joint goals previously adopted by the European Union. The combination of joint goals and indicators in social inclusion policy, in conjunction with the process of continual monitoring and evaluation of applications, has had a positive effect on the newer EU member states, which did not have a tradition of separate social inclusion policies in the past. The contribution of the EU in this respect has been of great importance as it not only gives priority to the poverty issue in these countries, but also establishes a basis on which to coordinate its reduction.

Judging by the comprehensiveness, in conjunction with the continual monitoring and implementation of European Commission recommendations of national social agendas, it can be argued that the effects of the ‘soft’ social mechanisms in the newer member states are more visible than those in the EU 15. Reasons that support this argument include: the discontinuation (in most cases) of previous social policy models and the need to replace them with those more effective; the use of the open method of coordination to search for best social policy practices; the effective use of financial and expert support provided by the EU; and finally, the need for quicker closing of negotiating chapters, which prompted a more immediate acceptance of social strategies offered by the European Union.

3. Social Policy Changes in Slovenia and Macedonia: Different Roads to the EU

Slovenia and Macedonia used to be part of one country, part of one political and social welfare system and part of the same ideology until 1991. After their independence, both countries placed identical political and strategic goals on their agendas: transition from a socialist to a functional market economy and association with the European Union. Fifteen years later, these countries are on the opposite economic, social and political spectrum. Slovenia has successfully managed to accomplish its political and strategic goals, while Macedonia continues in its attempt to reach them.

This part of the article intends to explore the factors that have contributed to these differences, seen through the ways and means used for social policy creation. The social policy changes are explored through an analysis of institutional, legislative and cognitive modifications evident in the social welfare systems of both countries. This part of the paper will also analyse the Europeanisation effects and differences in these countries and endeavor to depict possible convergence or disparity with European social standards and norms. In presenting the evidence in the selected case studies, this article will use the macro level approach in its analysis of social policy changes.

3.1 Institutional Changes

An analysis of the institutional setup for social policy administration in Slovenia shows that the following factors have contributed to its current structure:

(i) Firstly, historical heritage in social policy, or path dependency, as defined by Pierson [10], had an important impact during the early years of Slovene independence (1991), when a few institutions were created and/or renewed because of the country's distant past corporatist tradition. The Agency for Pension Insurance, the Agency for Health Insurance and the Agency for Unemployment were formed in the early 90's, which indicated that 'schemes for compulsory social insurance are an important part of the Slovene social welfare system' [11]. These agencies, while financed through the budget, are relatively autonomous, as they are not part of the public administration system.

(ii) The EU played another important role in influencing the institutional shape of social policy administration. Its directives and recommendations in the social sphere contributed to the following institutional innovations:

- Formation of the Public Guarantee Fund (1997) for workers in cases of employer insolvency. This Fund was formed in accordance with Council Directive 80/987/EEC, Article 5.
- Changes within the Ministry of Labour, Family and Social Affairs, which included: a) formation of a sector for the coordination of European Social Fund activities (2001); b) extension of the sector for international cooperation with the Sector for European Integration and c) the renaming of the Sectors into Directorates for easier coordination with the EU.
- Formation of a Central Unit within the Centre for Social Work for the purpose of coordinating social security schemes (family transfers, maternity leave, parental leave, etc.).
- Renaming the Sector for gender policy into the Sector for equal opportunity (2001), according to the EU directive addressing gender equality;
- Formation of the Agency for Invalids, although the activities of this body were transferred to the Ministries of Health and Labour as of 2004.

Institutionally speaking, changes in Macedonia show a rather different picture than that of Slovenia. The pace of institutional change has been relatively slow and initiated through different channels:

(i) International financial organisations, mainly the World Bank, are a major catalyst for institutional change. This was initiated in the late 90's and continues to the present day. Hence, certain World Bank projects in the social sector led to a number of institutional arrangements such as:

- Enhancement of technical equipment and human resources in the Agency for Employment (on the national and local level), in order to implement more active labour market policies (through technical assistance for institutional building to support employment policies);
- Enlargement of the institutional capacity in order to administer the new pension reform (through the SPIL-social protection implementation loan), which indirectly contributed to the formation of the new Agency for Supervision of Fully-Funded Pension Insurance (MAPAS). Through the initiation of pension reform, new private pension providers were established, whereby the responsibilities and pension contributions directed to the Public Fund for Pension and Disability Insurance were decreased.

(ii) The EU has only recently (2003-onwards) initiated projects, which have had some impact in the social sphere, and their institutional impact has been indirect:

- The PHARE Project (Institutional capacity building and development for social protection) contributed to the practical initiation of the previously planned pluralism in the social service provision, through supporting the partnership between the Ministry and the NGO sector. It also involved support for improvement to the management techniques undertaken by the Ministry of Labour and Social Policy.
- The CARDS Project (Technical Assistance to Institution Building in Support of Employment Policy, 2003-2005) delivered training to staff of relevant ministries and social partners' organisations on the carrying out of monitoring and evaluation activities on the implementation of NAP 2004-2005 measures and on employment policy development.
- In order to successfully coordinate EU issues, the Ministry of Labour and Social Policy enlarged its international department to include the Unit for European Integration.

In general, it can be argued that the institutional changes initiated in the two countries differ in terms of their direction and continuity of change. In Slovenia, there has been an overall preservation of the public institutions that existed in the past. Their restructuring complied with the overall socio-economic and political needs, and also with the requirements for European integration. In Macedonia, due to the more direct influence of international financial institutions and a limited budget, the public system of social welfare institutions that existed in the past is in a period of decline and its functions are slowly being transferred to the (as yet not developed) private and (fragile) non-governmental sector. European Union institutional requirements are at most, modestly present and their implementation would probably prove to be a time consuming and challenging process.

3.2 Legislative Changes

Through an analysis of the legislative changes in Slovenia since its independence, three different periods can be detected. The first period stretches from 1991-1996, when new laws were introduced which envisaged social policy be carried out through active measures, through cooperation with other sectors and also maintained a high standard of social security for its citizens. The second period, from 1996-2004 can be associated with the social policy changes that were undertaken as a necessary precondition for EU accession. The final period, from 2004 onwards, shows the creation and administration of social policy as both a result of national and European priorities and needs.

During the initial period, few legislative acts were created and adopted: the Law on Social Welfare in 1992 introduced the private and NGO sector as providers of services and also introduced an increase in social assistance benefits, the introduction of the guaranteed minimum wage (1995); active employment measures that included the implementation of Eurostat statistical standards, as well as the first analysis on the poverty issue. It is important to mention that it was this period that the first efforts to reform the social security system through legislative changes to the pension and health system (in 1992, 1993) were undertaken. Although legislative changes, especially in terms of the pension system, were not ambitious or radical, "the new government expected that they will contribute towards stabilization of the pension system" [12]. However, due to extensive and constructive public debate, the parametric reforms to the pension system did not occur until much later.

During the second period of social policy development in Slovenia, much legislative activity addressed EU requirements. Therefore, as a result of social *acquis* and the “soft” EU mechanisms, the following legislative changes and innovations took place:

- Changes to the Employment Relations Act (2002), which incorporated EU directives on labour law;
- Changes to Family Law and family benefits (2001), which incorporated a directive for parental leave and other complementary rights that regulate equality between genders.
- A new Law addressing health and safety at work (2001), which partially incorporated EU directives in this area
- Joint Assessment of the Employment Policy (2000)
- National Action Plan for Employment (1999)
- Joint Memorandum on Social Inclusion (2003)

Apart from the EU requirements, Slovenia expanded its social legislation through the creation and adoption of the National Strategy for Social Protection (2000) and the Programme for the Fight against Poverty and Social Exclusion (2001). Slovenia was assessed by the EU as to what level the criteria stipulated in these two documents had been fulfilled. What is particularly unique about this period is the adoption of pension reform (2001), whose final form came about as a result of a multitude of factors. Despite efforts by the World Bank to convey their preferred parametric reform offered in all other post-socialist countries, Slovenia accepted a fully-funded, modified pension model based on a voluntary second pillar. This was as a result of public pressure initiated by the trade unions, and was also supported by the EU PHARE project at the time. However, in 1996 there had been strong criticism expressed against the original reform suggested by the World Bank. The influence of this particular PHARE project was illustrated as being ‘in the right place, at the right time’ [12].

The third period of social policy development is a continuation of the previously undertaken steps. A new National Action Plan for Employment was created for the period of 2004-06, as well as the National Action Plan on Social Inclusion (2004). The change on the political scene as of October 2004 (when a right-oriented Slovene democratic party won the elections) may pose a potential danger to the previously undertaken reforms in social policy, as there is present fear among certain national players that the new government will ‘strengthen the principle of subsidiarity and give a bigger role to the NGO sector [11], and that ‘it will abandon its place in the tripartite dialogue, by providing more ‘opportunity for dialogue between workers and employers’ [13].

Macedonian legislative changes to social policy after 1991 were initiated rather late, and their pace was sporadic and ad hoc. The reasons for legislative changes, on the one hand, addressed the reform requirements conditioned by the international financial institutions (IFI’s) but on the other, witnessed a discontinuation of the strategic goals when power shifted between the different political parties.

The first more concrete and important legislative modifications were undertaken as of 1996. They included: initiation of the Labour Force Survey (1996), changes to the Labour Law (1997), which abolished labour market restrictions, adoption of the Unemployment Insurance Act (1997), which introduced more active employment measures, the Social Protection Law (1997) which introduced but did not specify the inclusion of new principles, such as public-private partnership, decentralization, etc., the Poverty Reduction Strategy Programme (2002), an Act for compulsory fully-funded pensions (2002).

These acts were created and adopted on the basis of financial and expert support from the World Bank and the IMF. Recommended reforms were adopted without prior public debate and critical evaluation. Although the Union of Trade Unions in Macedonia opposed this recommended paradigmatic pension reform, their disagreement came ‘too little, too late’. According to the trade union representative ‘the government publicly announced that the acceptance for pension reform is an obligation imposed by the FESAL 2 arrangement and we have to accept it as such’ [14]. This same

trade union representative also noted that the ‘government never takes any opposition from the trade union into serious consideration’.

Legislative changes introduced from 1996 until 2003 primarily focused on: (i) the introduction of selectivism and targeting in social service delivery; (ii) an emphasis on workfare instead of welfare; (iii) the introduction of private providers of social insurance; (iv) an increased emphasis on individual responsibility in place of solidarity as the main principle of social service delivery.

These reforms were not connected to either the social policy tradition of Macedonia, nor to the actual strengths and needs of the labour market. Social policy creation monopolized by the IFI’s may jeopardize achieving EU social standards in a prompt and appropriate manner. Hence, the Macedonian example, at least for the period following up to 2003, exemplified the Vaughan-Whitehead remark, that some of the EU candidate countries “implement neo-liberal reforms in a much more radical way than could be expected from a European state, thereby risking the future of Social Europe in the newly enlarged European Union [15].

The lack of EU influence in Macedonia until 2003 can be attributed to both the strong presence of international financial institutions, but also to the fact that Macedonia applied to the EU in March 2004, after which period more concrete EU instruments were made available in the social sphere. The EU CARDS programme contributed to the creation of the first National Action Plan for Employment. This instrument also plans to support asylum and migration issues through the preparation of a Law on asylum and the set up of shelters for asylum seekers.

However, an analysis of the Macedonian National Strategy for EU Integration gives a worrying picture of the country’s social policy priorities in terms of EU association. The main principles of social protection reform according to this strategy include limiting the egalitarianity of the social protection system and placing the responsibility for social protection on each individual and family. Plans for this subsidiary role of the state include a reduction in social transfers, the introduction of more rigid criteria for social protection and the limitations of social services to those most in need [16]. These principles are in sharp contrast with the EU general tendencies for maintaining high levels of social expenditure and universal access to social services.

3.3 Cognitive Changes

In regards to the changes in terminology, agendas and discourse within the social policy of the two countries, a different picture, once again, emerges. In Slovenia these changes are more visible, as they can be noted through the daily practice and activities of all parties involved in social policy administration. The increased use of national action plans in areas of employment and social inclusion, the logic of the European Employment Strategy and the practice of the open method of coordination have impacted on the behavior and language of the administrators, academics, trade unions and NGO representatives in Slovenia. There is very little evidence of such behaviour and language in Macedonia, where even on the ministry level, it is not evident, let alone the evidence of some knowledge of current European values and approaches that are used as a basis for discourse by the social players.

In order to make a more concrete comparison of Slovene and Macedonian social policy orientation towards EU preferred goals and standards, an assessment of their policy goals, instruments and effects will be shown to detect any possible trends towards convergence or disparity.

The table and graphs show that the two countries demonstrate a different pattern of adopted policies and instruments, as well as different outcomes in the social sphere. Slovenia clearly illustrates a significant level of convergence with the preferred EU standards, while Macedonia represents a case of obvious disparity with the EU norms. Despite the fact that Macedonia is only at the start of negotiations with the EU and has a very different economic potential than that of Slovenia, it may be argued that its prospective association with the EU may contribute to the diversity of social policy approaches and models existing within the EU. If that is the case, the prospects for a unified European Social Model will even be more hindered.

Convergence of policies	EU goals	Slovenia	Macedonia
Employment	(1) Achievement of full employment (2) Improvement to quality and productivity of work (3) Strengthening of social cohesion and inclusion	(1) Balanced socio-economic development (2) Increase in employment; (3) Increase in the living standard	(1) Active labour market policy (2) Reduction in unemployment; (3) Reduction of disparities among unemployed
Social inclusion	(1) Facilitating participation in employment (2) Prevention of social exclusion (3) Assistance to vulnerable groups; (4) Mobilisation of all relevant players	Adoption of all EU goals within the National action plan for social inclusion.	Declarative statements for social inclusion with no concrete action towards EU goals
Pensions	(1) Adequacy of pensions; (2) Financial sustainability of the pension systems; (3) Modernisation according to the needs of the economy, society and individuals	Reform of the pension system encompasses the EU goals in pension policy	Reform of the pension system is not compatible with the EU goals in pension policy

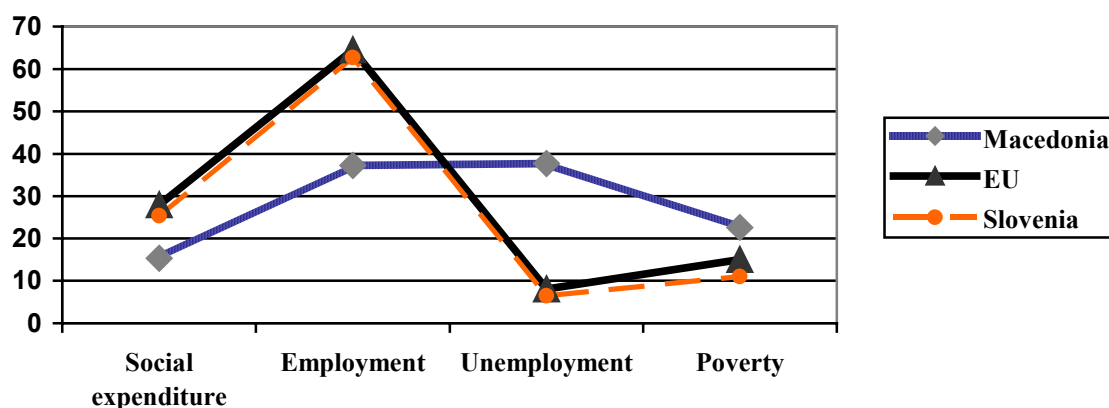
Table 6 Convergence of social policy goals

EU instruments	Slovenia	Macedonia
Social Acquis	Adopted	Initiated
Open Method of Coordination	Participating in OMC	Not participating in OMC
Social (structural) indicators	Applied	Not applied
Joint Assessment of the Employment Policy	Adopted	Not prepared
National Action Plan for Employment	Adopted	Prepared
Joint Memorandums for Social Inclusion	Adopted	Not prepared
National Action Plan for Social Inclusion	Adopted	Not prepared

Table 7 Convergence with the EU instruments

Indicators	EU 15	Slovenia	Macedonia
Social expenditures as % of GDP	28.0 % (2002)	25.4% (2002)	15.3% (2003)
Employment rate	64.5% (2003)	62.7% (2002)	37.3% (2004)
Unemployment rate	8.1% (2003)	6.5 % (2003)	37.7% (2004)
Poverty rate	15 % (2001)	11% (2002)	22,6% (2000)

Table 8 Convergence of effects in the social sphere



Graph 1 Degree of divergence (deviation) from the EU average in the social sphere

Dimension	Slovenia			Macedonia		
	Level of convergence			Level of convergence		
	High	Middle	Low	High	Middle	Low
Convergence of policies		√		√		
Convergence of instruments	√					√
Convergence of Effects	√					√

Table 9 Level of convergence of national social policies, instruments and effects with the EU level

4. Conclusions

The assessment of the social policy changes in the two analysed case studies show different Europeanisation effects. In Slovenia, a higher degree of Europeanisation is found, which can not only be observed on the legislative, institutional and cognitive level, but also on the levels of output. However, the contribution made by the EU for social policy changes in Slovenia is only one of the factors for this.

Other important elements that affected the overall social policy trajectory and setup include: (i) a prompt but incremental approach towards social policy reforms; (ii) the preservice and continuation of past social policy experiences and achievements; (iii) the strong influence of the trade union accompanied by a cooperative and responsible role of the state; (iv) the lack of influence from international financial institutions because of high economic growth and capacity and finally, (v) the continued and strategic pace of reform. Hence, the EU contribution was only one of support for achievements previously initiated and the ongoing reforms.

In Macedonia, the Europeanisation effects are not that evident, but this is not only because the country has recently gained EU candidate status. It may be argued that because of the strong and wide impact of international financial organisations, social policy in Macedonia has been reformed according to the neo-liberal preferences, who are a minority within the EU. Apart from the enormous role IFIs have played in shaping social policy, in Macedonia there is also: (i) the absence of national opposition to recommended neo-liberal reforms; (ii) the absence of a national consensus on an ideological trajectory for social policy and minimum social standards; (iii) weak trade union organisation; (iv) the declarative, yet not formally inclusive and cooperative role of the state for working with social partners and (v) low economic growth and capacity of the country. All this, in conjunction with

minimal EU presence in social policy reform in the past, has contributed to the creation of a model, which can be described as a neo-liberal rudimentary one. Rudimentarity can be explained through the non-existence or the minimal capacity of the relevant elements required for the functioning of a neo-liberal model, such as a strong private sector, high economic capacity, a functioning tax system, etc. Thus, Macedonian social policy is in considerable disparity with the current EU social standards and preferences.

The inclusion of countries like Macedonia within the European Union will probably contribute to a greater diversity of social policy models, endangering the deepening role of the EU in social policy. Therefore, to preserve and enhance its social policy effects within the member and candidate countries, the EU has to root its role in social policy through a more direct and concrete social policy prescription based on its legislative and institutional requirements. In that way, the EU will be able to become a competitive player in the globalised environment, and consequently, indirectly improve future social policy convergence within its borders.

Acknowledgments

I want to thank my supervisors Bob Deacon and Paul Stubbs for their continuous support and encouragement throughout my PhD research, of which this article is a small part.

- Bob Deacon, University of Sheffield, Dept. of Sociological Studies, Elmfield, Northumberland Road, Sheffield S10 2TU, B.Deacon@sheffield.ac.uk
- Dr Paul STUBBS, Senior Research Fellow, The Institute of Economics, Zagreb, pstubbs@eizg.hr

References

- 1 Ioakimidis, P.C. 'The Europeanisation of Greece: An Overall Assessment' in Featherstone, K. and Kazamias G. (eds.) *Europeanisation and the Southern Periphery*, London: Frank Cass, 2001
- 2 Radaelli, C. M. 'Wither Europeanisation? Concept Stretching and Substantive Change', paper presented at the Political Studies Association Annual Conference (Apr.), London, 2000
- 3 Guillén, A. and Palier, B. "Does Europe matter? Accession to EU and Social Policy Developments in Recent and New Member States" in *Journal of European Social Policy*; 2004, Vol. 14(3): 203–209.
- 4 Lendvai, N. "The weakest link? EU accession and enlargement dialoguing EU and post-communist social policy" in *Journal of European Social Policy*, 2004, Vol. 14(3): 319-333.
- 5 Falkner, G. *EU Social Policy in the 90s: Towards a Corporatist Policy Community*, London: Routledge, 1998
- 6 Bertozzi, F. and Bonoli, G. "Europeanisation and the convergence of national social and employment policies: What can the Open Method of Coordination achieve?", Paper prepared for the workshop "Europeanisation of national political institutions", ECPR Joint-session, Turin, 22-27 March 2002.
- 7 Börzel, T. and Risse, T. "Conceptualizing the Domestic Impact of Europe" in Featherstone, K. and Radaelli, C. (eds) *The Politics of Europeanization*, Oxford: Oxford University Press, 2003
- 8 Kleinman, M. *A European welfare state? European Union social policy in context*, Basingstoke: Palgrave, 2002
- 9 European Commission, *Draft Joint Employment Report 2004/05*, SEC(67), Brussels, 2005
- 10 Pierson, P. 'Increasing Return, Path Dependence and the study of Politics' in *American Political Science Review* 94, 2, 2000, pp. 251-267.
- 11 Kolaric, Z. full-time professor at the Social Science Faculty, University of Ljubljana, Excerpt from an interview carried out for the purpose of the PhD entitled: *Influence and Impact of Europeanisation on the Making of Social Policies in Slovenia and Macedonia*, January 2005.
- 12 Stanovnik, T. "The Political Economy of Pension Reform in Slovenia" in Fultz, E. (ed.) *Pension Reform in Central and Eastern Europe*, Vol.2, Budapest: International Labour Organization, 2002

- 13 Mazalin, B. Chairman of the Confederation of trade unions 90, Slovenia, excerpt from an interview carried out for the purpose of the PhD entitled: Influence and Impact of Europeanisation on the Making of Social Policies in Slovenia and Macedonia, January 2005
- 14 Jankulovska, L. Representative of the Union of trade unions of Macedonia, Excerpt from an interview carried out for the purpose of the PhD entitled: Influence and Impact of Europeanisation on the Making of Social Policies in Slovenia and Macedonia, January 2005
- 15 Vaughan-Whitehead, D. EU Enlargement versus Social Europe: The Uncertain Future of the European Social Model, Cheltenham: Edward Edgar Publishing, 2003
- 16 Government of the R. Macedonia, National Strategy for Integration of R. Macedonia in the European Union, Skopje: Government of the R.M., 2004.

Identities and Perceptions: The Changing Status of English as a Lingua Franca

Sara Hannam¹, Ann-Marie Bathmaker²

¹Lecturer, English Language Unit, City College, Tsimiski 13, Thessaloniki, 54624, Greece, hannam@city.academic.gr

²Lecturer, Department of Educational Studies, University of Sheffield, 388 Glossop Road, Sheffield, S10 2JA, UK, a.m.bathmaker@sheffield.ac.uk

This paper outlines the first phase of part-time Doctoral Research in the area of English as a Lingua Franca, carried out by Sara Hannam and supervised by Dr Ann-Marie Bathmaker, which will be applied in Serbia and Bulgaria in 2007/8. The rationale informing the initial choice of topic is considered, as well as the overall theoretical framework developed for the field work and analyses which was arrived at through an extensive and multi-disciplinary literature review. The key concepts and terms considered necessary in this project are also presented. I argue that English as a Lingua Franca (ELF) occupies a complex and contradictory status in the Balkan region and that its existence and use could potentially be perceived in an number of ways, both positive and negative. I also posit that the relationship the individual has with ELF is directly related to issues of identity and the socio-political background that forms the current Balkan landscape. A detailed understanding of the discursive construction of the ‘Balkans’ throughout history, notably by the Western Academy and Media, is presented as essential, as well as the continued rigour of the researcher in attempting to deconstruct and challenge such images to gain a deeper understanding of the region and countries under review

Keywords

Balkanism, Ethno-linguistics, Identity, Social Perception.

1. Developing English(es)

Until very recently, those learning or using English as a second language in contexts outside countries where English is the first language, were generally referred to as EFL learners/users (English as a Foreign Language). This term was premised on the assumption that interactions in this setting were likely to take place between a Native Speaker (N-S) of the language and a Non-Native Speaker (NN-S), the former usually being the teacher and the latter the student. The language model that was aspired to, in terms of structure and pronunciation, was what is known as the *prestige* version of N-S English (often referred to as Queen’s English or Standard British English in the UK and General American or Standard American English in the USA). This was presented as a single variety, at the expense of all other regional versions of the language that might exist in the source countries themselves e.g. UK, USA, Australia etc. This was also the model that course books and learning materials tended to utilize, and that language teachers, testers and learners alike thought of as ‘correct’. It is true to say that this view is still prevalent in English Language Teaching (ELT) today and informs many of the pedagogical decisions made at different levels of the profession, as well as the all-important area of how language skills are assessed and categorized.

However, the historical and economic legacy of England’s Colonial past and the dominant position of the USA in the contemporary world have ensured that “for the first time in the history of the English language, second language speakers outnumber those for whom it is the mother tongue” (p. 1 [1]). The figures regarding actual usage speak for themselves and it is clear that the Native Speaker of the English language is now in the minority – 1996 statistics claim that there are four N-NSs for every one

N-S in the world [2]. Many situations where English is used nowadays no longer involve any N-Ss (the research cited above claims 80%) and the English language is therefore evolving in a variety of fascinating ways, mirroring the diversity of users and uses. The South East European Research Centre (SEERC) itself, and the Doctoral Conference from which this paper emerged, are a case in point in this respect as N-NSs are the majority in this setting. It is therefore the case that different English(es) *are* already the norm and we should be moving forward into an era where this is both acceptable and something to be celebrated. Despite the ELT profession having been aware of the global role of English for quite some time, this “realization... has not so far led to any radical reconceptualization of that English” (p 212 [3]). In other words, the teaching and learning of English, as well as the way it is perceived in the wider community, is often based on outdated assumptions about who owns it, uses it and is ultimately responsible for its future development.

2. What Is ELF?

The term English as a Lingua Franca (ELF), also sometimes referred to as English as an International Language (EIL), is a recent development in the field of Theoretical and Applied Linguistics, although it has been around as a phenomenon for many years. The decision to abandon the name EFL by some researchers and teachers, and to replace it with ELF, was a conscious one. It is based on the change of goals in language learning and/or use as outlined above and influenced by a pragmatic change in the way in which English is used globally. In other words the terms highlights the fact that ELF is a variety spoken exclusively by N-NSs of the language and, in the vast majority of situations, is used as a means of communication and “as a contact language among speakers who come from different first languages and cultural backgrounds” (p. 9 [4]). Descriptions of ELF, however, are a recent development and are not complete – they have focused mostly on analyses of the spoken language and have found, among other things, that certain items of lexicogrammar are repeatedly used in ELF interactions despite them being considered ‘wrong’ in an EFL setting. For example, as pointed out by Jenkins [ibid] the third person ‘s’ in *he says* is often omitted, extra prepositions might be added such as *to discuss about*, and question tags (such as *isn’t it*) are over-used. Other examples might include confusing the which/who pronoun and using ‘which’ to refer to people or ‘who’ to objects. All these grammar/vocabulary points can be found in EFL coursebooks as ‘mistakes’ in need of repair and correction. However, in ELF interactions, such uses do not affect comprehension between speakers and are considered perfectly normal. Why then do powerful institutions, such as publishers and language testing agencies, “continue trying both to exert a strong influence in favour of Native Speaker norms and to promote negative attitudes towards ELF norms among the very Non-Native Speakers who stand to benefit from them?” [ibid]. This question will form one of the central concerns of the thesis.

At an ideological level, the concept of ELF is concerned with redressing the balance and levelling the playing field for N-NSs - as teachers, learners or speakers. In other words, if an ELF approach to using or learning the language is adopted, a far more achievable range of points can be considered ‘essential’ for successful communication – they are likely to be based only on what might cause problems with understanding, rather than what is ‘correct’ according to the N-S benchmark. This more inclusive model suggests both equal opportunity to access as well as increased chance of success in a wide range of situations such as employment and academia. NN-Ss are well documented as having greater problems proving their proficiency in the language or getting articles published in English speaking journals, to name but a few examples, unless they are prepared to conform to N-S norms laid out by journals. Additionally, ELF parameters recognize that language is connected to identity and is an important part of what it means to be human. Respecting individual rights and therefore uses of ELF is fundamental in ensuring its successful development and the elimination of outdated patterns of discrimination and practice. In other words, ELF parameters allow us to deconstruct the way in which language has been used to acquire privilege and power and offers a chance to readdress some of the damage of the past that has been caused by colonial practices, an integral part of which was the imposition of a particular cultural reality (and language) on local populations [5]. It suggests a more critically aware pedagogy which is in tune with *post*-colonial developments or, as pointed out by Kramsch: “It is the responsibility of applied linguists, as

researchers and as language educators, to openly show the connection between language problems and the larger historical and geopolitical conditions that have brought them about, and to re-frame the problems accordingly” (p 562 [6]). The fact that the field of Linguistics is currently involved in an intense debate over new and existing pedagogy, with a significant amount of resistance to change being prevalent, provides a very fruitful area for research.

3. ELF in the Balkans – Why Now?

Bearing the above background discussion in mind, at this stage the specific PhD research has four aims which could be categorized as follows:

- To investigate the changing status of ELF at the beginning of the 21st Century
- To investigate the use of the English Language in shaping and creating identities in countries where English is not the first language
- To investigate the hypothesis that language is a site upon which anxieties regarding changing status(es) are negotiated and debated
- To interrogate the notion that teaching/learning ELF is a neutral activity

These four general aims will be pursued in the context of the Balkan region as a geographical location, and specifically in Serbia and Bulgaria through a series of more specific objectives as follows:

- To investigate the changing environment in relation to the use of ELF in Bulgaria and Serbia as post-Communist countries, and Serbia as a post-conflict zone
- To investigate how English is perceived in these settings by ELT practitioners and professionals and to assess differences and similarities across the Balkan landscape
- To investigate what other forces are in operation that influence how English is perceived e.g. The British Council, The Fulbright Commission, the Armed Forces, The European Union
- To interrogate the validity of the widely-held belief that the dominance of English as the second language of the region is an overwhelmingly positive development

The reasons for choosing the Balkans as a research environment are too numerous to mention, but I will outline the most pertinent ones at this stage in the research. Primarily the initial choice was based on the fact that English has quite suddenly and quickly been identified and imposed as the second language for all (through the public and private education system), and in the last 10 years or so has overwhelmingly been presented as a positive development for the inhabitants of the countries in the region. This is actively promoted by external agencies such as the British Council through their work in the area, and promotion of language tests which derive from the UK (and USA). It is now the case that almost all employment and development opportunities carry with them the assumption of a knowledge of English and a specific level of certification or ‘proof’ of knowledge.

There are a number of important considerations in relation to this development. Firstly, there is a need for recognition of the role that English plays in the global business world and that it is often the Lingua Franca in settings where educational/employment opportunities are to be found. However, it is also true to say that the exporting of language learning and testing is a very profitable industry and its potential is currently being exploited through intensive advertising of the necessity for learning and testing (based on EFL models) [7]. Secondly, it is often identified as a ‘neutral’ language in a region where ethno-linguistic identity has been an issue of tension and conflict for a number of years. During the Ottoman period and after there is significant evidence to suggest that the inhabitants of the region that is now termed the Balkans were multi-lingual [8]. Language and ethnic identity were not viewed in the same way that they are today, and were not the source of severe splits between communities that they have become in recent history. In a contemporary setting, as the local inhabitants have lost the ability to communicate in their neighbour’s tongue (often enshrined in language laws elevating the use

of specific languages as ‘official’) [9], the use of English has grown and continues to flourish as a ‘necessity’ to be successful in today’s world. This creates a profound tension worthy of investigation.

Another interesting area of development following the cessation of the NATO bombing of Kosova and Serbia and the installing of troops for peacekeeping purposes, is what has become termed the Peacekeeping English Project (PEP). This is driven forward by the British Council on behalf of the British Foreign and Commonwealth Office. The aim of the project is, according to one internal publication:

To promote respect for civilian democratic government and practices...and to enhance understanding of the UK’s democratic values and processes, and work in partnership with other countries to strengthen good governance and human rights [online: 10].

There are two interesting issues arising from this development. The first is the fact that the troops stationed in the Balkans (from various countries) rarely speak the local languages and therefore, once again, English is the Lingua Franca of communication between them and the local environment, although often not their own personal first language. The second is the flourishing of a new sector for ELT which is constantly in search of new markets and the decision by MacMillan, one of the largest publishers, to produce a series of coursebooks called *Campaign* for the purposes of teaching this ‘variety’ of English. The coursebook is a somewhat unsettling combination of the language of social interaction and engagement with the local community in pursuit of peace in the region, along with, for example, instruction on the names of parts of a rifle in English for assembly purposes – the latter perhaps reveals the other face of the armed forces in the area and results from the reality of the armed intervention that has formed and continues to form a major part of their presence. Finally, and perhaps most interestingly for this PhD research, there is a taken for granted assumption in the above quotation that democratic processes are a) in place in the UK and are b) not in place in the Balkans – again this assumed contrast is worthy of further exploration.

One more important factor in the development of ELF in the region is the impact of the European Union (EU). Since the breakup of Yugoslavia, many of the countries of Former-Yugoslavia, as well as surrounding countries, are anxious to become part of the EU. At the date of writing, this process has not resulted in acceptance for many of the countries, with the exception of Greece and Cyprus, with Bulgaria and Romania promised membership in 2007 if they can address specific issues such as corruption and immigration. There are two issues arising from this reality that are relevant to this PhD research. Firstly, the language of communication in the setting of decision making and application to become a member of the EU is inevitably English. This means that each country that wishes to become part of the EU must strive to present itself in the best possible light not in its native tongue, but in ELF. Secondly, and perhaps more importantly, the issue of the dominance of English within the EU has become of major significance. Despite strong feelings from many of the members that they should have the right to communicate in their own language, and the promise from the EU that all languages are equal, the reality is often quite different. Recent research demonstrates that English is dominant in a whole range of situations and functions as a symbol of power relations as the EU moves towards a single-language state [11]. This development parallels (at the time of writing) the recent negative comments made by the British Government regarding the expected immigrants from Romania and Bulgaria post-2007 who, they have made clear, will not be greeted with the same welcome as those from other Eastern European countries, such as Poland, who came before them.

When taking the above considerations into account, it is possible to conclude that English in the region might be seen in three potentially different ways. It might be identified as a liberating force, as mentioned above, in a region where language and ethnicity have been the source of much conflict historically. Alternatively, English could be viewed as an oppressive force, which is attached to memories of NATO intervention or the notion of ‘Linguistic Imperialism’ as outlined by Phillipson, or the dominance of English as a form of Cultural and Economic Capital [7]. A third more negotiated position might be that English is an oppressive force that also has a liberating potential, particularly if it is used to deconstruct and critique the structures that have led it to become so dominant. In other words, it can also be the language of resistance to some of the forces outlined above and could be used

as a means of articulating a more meaningful identity to the outside world than that which is currently being projected.

4. Research in the Balkans – Defining the Parameters

The term “Balkans” is not used without recognizing its socio-historical complexity – it is utilized in the interests of enabling myself as researcher to confront the ramifications of the regional history in all its enormity rather than to move away from its negative signification which would be another possible option. Iordanova (p 6 [12]) argues that referring to the Balkans “as a common denominator when referring to this diverse and complex region...allows (us) to name and critique important transnational issues the often remain neglected when the exploration is limited to individual countries”. It is for this reason that in this particular project, the region has been termed as such, the future plan being to concentrate on two different countries within this region, namely Serbia and Bulgaria. The choice of these two countries is based on a) my connection within these two countries with the ELT environment which gives rise to a rich potential interview source b) the fact that both countries are what could be termed ‘post-Communist’, although each arguably experienced the Communist period in different ways and have coped in different ways since the fall of Communism and c) Serbia is also a country that could be termed ‘post-conflict’ which enables a consideration of the notion of ‘Peacekeeping English’ as defined above.

My own nationality and geographical location (British, living in Greece for 11 years) has also been a complex area of consideration, in the sense that the project requires me to situate myself. In carrying out the literature review to define the parameters of researching the Balkans, I drew extensively on research and writings from Balkan academics who are living outside the region in Western countries and writing in English, although I will also be investigating local educational sources written from within the region. I am, therefore, an ‘outsider’ in the region which will need to be fully taken on board in the research process. I am attempting to critique the Western view of the Balkans, as opposed to the Balkans or the Balkan people themselves. It is my supposition that the way in which the English language has been imposed in the region, along with a specific economic agenda, will allow me a window into this complex representation, which interplays with historical and cultural developments.

Having read extensively from the more critical sources on Balkan developments that exist in academia today (Mazowar and Todorova being two of the most influential) during the preliminary literature review for this PhD, as well as other sources that allow access to ways of ‘reading’ the region (Said’s notion of ‘Orientalism’, Bakic-Hayden’s notion of ‘Nesting-Orientalisms’, Bhaba’s notion of ‘Liminality’), but not (unfortunately) having the time or scope in this paper to explore all these ideas thoroughly, the theoretical position that I have chosen to take in relation to my research in the Balkans will be briefly described below. Todorova defines a way of understanding the Balkans termed *Balkanism* in which she rejects the Saidian notion of Orientalism as applicable to the region. Rather, she believes that the lack of *victimhood* found among the Balkan populations demonstrates that people do not see themselves as post-colonial subjects at all. Orientalism is interested in the tension between the West and the Orient, whereas Balkanism is interested in the tensions between different strata within Europe, in which *Balkan* has been designated an inferior position and has become “Europe’s shadow, the structurally despised alter-ego, the dark side within” (p 18 [8]). This leads to a very different relationship with the concept of Other, and Todorova concludes that “this in-betweenness of the Balkans, their transitional character, could have made them simply an incomplete other; instead they are constructed not as other but as incomplete self” (Ibid). This fusion of critical theory and more psychoanalytical considerations is an interesting paradigm to work with, particularly in relation to the interplay of the local languages of the region and ELF.

Whilst I believe that Todorova’s analysis has much to offer the researcher interested in the region and have drawn extensively on her work, I also feel that it has some limitations, perhaps best expressed by Bjelić and Savić [13] who consider the wealth of possible theories available for analyzing the Balkans and draw a number of more contemporary conclusions. Bjelić touches on the nature of Balkanism as a critical field of enquiry and makes the valid point that unlike Orientalism which had behind it a huge

intellectual and theorizing machinery “only very recently, with the disintegration of the former Yugoslavia, was “expert” knowledge introduced in Balkanism” (p 7 [13]). Unlike Orientalism which was developed within the Occident and its halls of learning, Balkanism has developed almost exclusively from *within* the Balkans itself. It could be argued that this is one of its strengths as it implies a more level playing field. However, although Balkan scholars might agree on the “general contours of Balkanism” it is also true that they “disagree on the proper methodology for its critical analysis” [Ibid]. Bjelić critiques many of the existing paradigms as they fail to draw attention to the link with a discourse of globalization (or Neoliberalism) and particularly the contradictory relationship with the European Union, of which language is one. As he argues: this contradiction represents “first a horizontal antagonism between the Balkan states and ethnic groups, in which each of them is a potential aggressor; secondly a vertical system of co-operation between each of these parties and the European Union” (p 8 [13]). This urges the commentator to see Balkan identity as a constantly evolving phenomena tied to socio-historical development in the here and now, and to take into consideration the economic aspirations of external players such as the UK and US, as well as the EU, in shaping the past, present and future in the region.

Bjelić (p. 15 [13]) concludes that the challenge for Balkan scholars (and I think it could be widened here to include other scholars who are interested in the region) is to recognize that this “universalized globalism....is ethnic too....(and) the challenge for Balkan scholars, despite their desire for a modern non-ethnic identity, remains one of recognizing how resistance to consumerist globalism is corollary to their resistance to nationalist myopia” [Ibid]. When considering how to proceed with this research given the above parameters, it is important to be aware of the limitations from both within and without the Balkans, from both Western and Balkan scholars. I would conclude that tying the analysis (and methodology) to the development of the global economic system under which we all live (including the Balkans) is essential in making sure that all structures, which have played a central role in the development of the region are recognized, and to ensure that the critical balance is correct. In that sense, as summed up by Bjelić: “Balkanism has bestowed on Balkan scholars an opportunity to mount representational resistance against the imperial depredations and shallowness of global culture” (p 19 [13]). I hope that it will be possible to extend this to scholars of other nationalities, such as myself, who wish to be part of this process. The history and development of the English Language in the region is the link through which I hope to explore these issues.

5. Future Directions: The Next Step

The next step in the PhD research is to carry out the fieldwork in Serbia, which will comprise a combination of individual and focus group interviews with the following target groups exploring the themes that have been raised in this paper:

- Serbian English Language Teachers
- N-S English Teachers living and working in Serbia
- Representative from the British Council “Peace Keeping English Projects”(based in the UK)
- Representative from local British Council in Belgrade
- Representative from Educational Services at Governmental level (responsible for the learning of foreign languages)

It may not prove possible to gain access to all of the latter three groups, and this process will be revisited depending on the results of the initial request process. The aim of involving a number of different perspectives is to ensure that the competing voices offer multiple parameters to consider in the analysis, thereby providing the richest and most balanced possible perspective on the changing role of ELF in Serbia. Through a series of questions which consider the various ways that English is currently being used in the region, and what function it has vis-à-vis the local languages, it is hoped that a deeper understanding can be gained on the three different possible ways that ELF in the region might be viewed (as outlined above). The first group of interviews are planned for mid-2007.

To conclude, it is hoped that this research will help to further explain the complex way in which language interplays with political development, and also with individual and national identity. This research will also add to the growing body of work on English as Lingua Franca, adding a new dimension in a number of new ways. Rather than considering the pragmatic use/usage of these new varieties, it will focus on the ethno-linguistic significance of such developments, in two countries that are not given a great deal of coverage overall in research and development terms in the field of ELT and often remain neglected as a source of enquiry. It is also predicted that the extra data will prove timely in the field of Applied Linguistics where the development of ELF and its ramifications are a constant source of debate and discussion.

References

- 1 Jenkins J, The phonology of English as an international language: new models, new norms, new goals. 2000. Oxford: Oxford University Press.
- 2 Kachru B, The paradigms of marginality. *World Englishes*. 1996; 15/3: 241-255.
- 3 Seidlhofer B, Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics*. 2004; 24: 209-239.
- 4 Jenkins J, The ABC of ELT....ELF. *IATEFL Issues*. (December-January). 2005: 9.
- 5 Pennycook A, English and the discourse of colonialism: the politics of language. 1998. London: Routledge.
- 6 Kramsch C, Post 9-11: foreign languages between knowledge and power *Applied Linguistics*. 2005; 26/4: 545-567.
- 7 Phillipson R, *Linguistic Imperialism*. 1992. Oxford: Oxford University Press.
- 8 Todorova M, *Imagining the Balkans*. 1997. New York. Oxford University Press.
- 9 Fraenkel E, Kramer C. *Language contact – language conflict*. 1993. New York: Peter Lang.
- 10 McIlwraith H, [Ed]. *Peacekeeping English project newsletter*. Issue 15/July. 2004. Accessed on-line at www.britishcouncil.org/english/pep/newsletter/current.htm [September 2005].
- 11 Phillipson R, *English only Europe? challenging language policy*. 2003. London: Routledge.
- 12 Iordanova D, *Cinema of flames: balkan film, culture and the media*. 2001. London: BFI Publishing.
- 13 Bjelić D I, Savić O. [Eds]. *Balkan as metaphor: between globalization and fragmentation*. 2005. Cambridge: Massachusetts: The MIT Press.

Appendix

List of acronyms

EFL	English as a Foreign Language
ELF	English as a Lingua Franca
EIL	English as an International Language
EU	European Union
ELT	English Language Teaching
N-S	Native Speaker
NN-S	Non-Native Speaker
PEP	Peacekeeping English Project
SEERC	South East European Research Centre

Integrated Border Management: Constructing Cross-Border Spaces

Despina Syrri

SEERC, 17 Mitropoleos Str, 54624 Thessaloniki, Greece, dsyrri@seerc.info

The paper aims to analyse critically the discourses, policies and practices of a set of diverse international development actors engaged in relationships and interactions of ‘integrated border management’, ‘cross border cooperation’ and the creation of ‘a common European space’. The paper also intends to explore the relationship between borders and governance in terms of the EU’s relationship with its own ‘borderlands’. While conducting fieldwork along the South East European euroregions under construction, nation-state territoriality as a container-limit for politics, is problematised. The question arises as to how the contradictions inherent to the ‘governmentality’ of EU enlargement appear in the South Eastern European borders, particularly as the imperatives of market expansion collide with the security goals of Schengen in the context of increased global securitisation.

Keywords

Borders, European Union, integrated border management, cross border cooperation, governance, South East Europe, international actors, security.

1. Introduction

The changing borders of the European Union, the ways in which they are created, maintained and reconfigured, reveal much about the Union’s identity as a transnational polity. The paper examines how processes related to the changing role of borders as barriers, bridges, membranes, resources and symbols of identity in contemporary Europe unfold, by attempting to look at what seems like the European project’s Janus face when seen from the South East European countries’ perspective. On the one side the “Integrated Border Management” recent security agenda translates mostly the historical legacy of coercive, imposed, sharply dividing borders, while on the other side, the development, policy making discourse of “Cross-Border Cooperation” speaks of more negotiated, consensual and ‘democratic’ ways of border change.

“Rather than accept the unilinear notion that territorial borders are becoming insignificant because of time-space compression or the increasing irrelevance of territory, it is suggested that a more fruitful line of enquiry is to empirically investigate territorializing and de-territorializing (or bordering and de-bordering) processes.” [1].

This paper explores the complex ways in which territory, both as space delineated by permeable yet excluding borders and as a site for specific projects of cross-border development, matters as a product of historical forces and multi-level practices. It looks at the EU’s border with the ‘Western Balkans’, as notions of ‘Euro-regions’ are exported South and therefore transformed from ‘laboratories of integration’ within Europe, into more complex experimental zones on and across Europe’s borderland, the construction of cross-border regions, in particular the negotiation and contestation entailed in development and security project interventions.

The paper aims to analyse critically the discourses, policies and practices of a set of diverse international development actors engaged in relationships and interactions of ‘Integrated Border Management’. It draws from recent literature on the merging of development and security, in the context of ‘new wars’, which, whilst territorially bounded, involve civilians and combatants alike, and have wider regional and global effects. Hence, a new regime has emerged comitted to ‘conflict

resolution and prevention' and 'social reconstruction' as a mode of governance in which both people and states are the objects of attention [2].

The formation of this 'new Europe' is a social process involving the interaction between power, coercion, citizenship, political representation and identity, constructed in sets of inter-state and state-locality relationships that are clearly manifested at the borders. An ethnography studying-up that which is taking place in and about the borderlands and frontier zones reveals much about the dialectic of fragmentation and integration that is currently reshaping the European continent. The process of the Western Balkans' integration into the European Union seems not to be only regulated by the official Stabilisation and Association Process (Sap), bilateral agreements between national governments and unilateral efforts by states. It is also heavily influenced by the presence and involvement of a multitude of actors; indicatively in border areas their work is focused on trade and transport facilitation (associated with the World Bank and the South East European Cooperation Initiative (SECI) [3], business promotion, security and migration issues. Depending on their agenda, the focus or interface differs, with the emphasis shifting from trafficking and organised crime to good governance and regulation of migrants. Shifting from an initial focus on transborder business activities and minorities issues, present activities are informed by the first difficulties of circumventing national governments' reactions and sensitivities, become interested in transnational crime and trafficking, customs reform, and crucially public sector reform. Thus the concept that haunts the EU euregios, namely 'integrated border management' becomes the governing principle in the Western Balkans too. The guarding of the borders against 'aliens' becomes a priority, thus the intensified policing of the EU borders-to-be is directed towards the prevention of the border crossings of non-European individuals.

"Border regions thus offer, at least potentially, always latently, always provisionally, a space for territorial politics beyond any one single nation, indeed beyond all nations. As to define the scale and scope of such a politics, recuperating the 'archaic' qualities of borders serves as a corrective to what we consider to be overtly 'presentist' and a-historical accounts of globalization and regionalization, recalling that current accounts of capital restructuring can mobilize only a finite repertoire of territorial logics and counter-logics, many of which have roots in pre-capitalist social relations. Border regions thus gesture towards a cautious return of time and the temporal, yet one (we hope) that is far removed from the totalizing historicisms of the last century. We suggest that it is on the canvas of this complex and contradictory European space-time that the topic of governance may now be framed concurrently as a cross-border regional question." (emphasis in original) [4].

2. Borders' policies in Europe

According to multi-level governance theory, advocates of a federal 'Europe of the Regions', allowing subnational groups more freedom of expression and influence, consider 'ever closer union' as the means of circumventing existing states without confrontations over territorial borders. Proponents of a looser 'Europe of the States' argue that intergovernmentalism is a way of preventing the emergence of an alternative to the existing inter-state system, namely of strong regional entities or inter-regional alliances. The potential threat posed to state sovereignty by the lack of correspondence between nation and state underlies the majority of borders' policies. The abolition of internal border controls came as a recognition of the weakening infrastructural power of the national state arising from the globalization of national economies and the minimising of state influence over production and financial markets. With the growing multinational control of the global economy, states lost much of their power of initiation, if not of regulation, thus disparities and conflicting interests developed between regions and economic sectors within states [5]. Significant elements of economic regulation would be shifted to the EU level as a way of influencing the continental strategies of multinational corporations, while states could attempt to undermine the delegitimizing potential of radical economic restructuring and mass unemployment by transposing economic responsibility upwards.

Trends as diverse as economic globalisation, international migration as well as fear of terrorism and organised crime pushed the efficient handling of borders to become an issue of political priority.

Modern, economy-oriented states are supposed to rely on flourishing trade and offer a comfortable degree of security to their citizens. The formula commonly chosen in combining these two objectives is that of 'integrated border management' (IBM), a term unknown until the mid-1990s, yet currently the keyword for innumerable reform projects around the world [6]. Born as an attempt to marry security concerns with trade facilitation, IBM's contents are as heterogeneous as the regions involved. There is a common understanding that IBM relates to 'lean government' approaches and that border procedures should be governed by modern economic strategies rather than slow bureaucratic structures, in short, should be tailor made for each case.

The implementation of this approach becomes a balancing act for an incomplete federation such as the EU, with its sensitive mix of a single external border and 25 separate legal/administrative systems. External borders and their crucial role for the smooth working of the Single Market, the Area of Freedom, Justice and Security and other basic objectives of the Union have long been recognised. 'Union' and (common) 'border' have formed an inseparable pair of notions since the early days of the Common Market and the EEC customs union in the 1960s. "Weak links in the border chain" and "the need for burden-sharing and solidarity" soon became keywords and phrases of an EU-wide debate.

2.1. Cross-Border Cooperation

The proactive EU borders' programme was a belated response to the work of a purely intergovernmental and consultative organization, the Council of Europe, on borders over a thirty year period, seeking to develop a triangular borders' policy involving cooperation between European institutions, national states and border regions. From the 1950s onwards a tradition of cross-border cooperation had been developing along the Rhine Axis on the Dutch, German, Belgian, French and Swiss borders [7]. Central to these attempts was the formation of the Association of European Border Regions which was later to become involved in monitoring the EU's INTERREG initiative. The Council of Europe continued to stimulate interest and research on border areas and cross-border cooperation [8], and promulgated a Framework Convention on Transfrontier Cooperation in 1981 which advanced new models and juridical principles for cross-border links, namely the Euroregions or Euregios. Yet this Convention had minimal practical effect as the EC/EU remained divided on the precise nature of the links between economic integration and national sovereignty.

The INTERREG funding initiative, specifically designated for internal and external border regions, had a dual purpose: to help integrate the economic space of the Community as a whole and to address the negative legacy of border areas, namely their isolation from the main centres of economic activity and decision-making, the separation of their commercial centres from their natural hinterland, their inferior infrastructure, their generally poor natural resources, and the specific difficulties arising from having different legal, administrative and social welfare systems, and often different languages and cultural traditions, from their contiguous region across the border [9]. The INTERREG programmes commenced in 1990, Interreg I lasting from 1990 to 1994. The second programme was launched in 1994 and ran until 1999. INTERREG III started in 2000 and runs until 2006.

Within Europe's cross-border regions, novel patterns of governance were supposed to address problems of taxation, pension transfers, accountability of decision making, management of funds, re-territorializing space in such a way as to reconcile the demands of 'borderless' liberal market efficiency with the vision of a democratic Europe of citizens. Many important, perhaps inherent, problems are yet to be confronted, and these are faced most starkly at Europe's periphery and external borders, such as the institutional design and institution building of these socially constructed entities, the political economy of cross border governance, as well as the complexities of local politics. The emergent forms of territorial governance within European regions established administratively since the early 1990s which aimed to span the former political frontiers separating EU member states, namely cross-border regions, have received financial impetus from the European Commission, and embody the hopes of becoming 'laboratories of European integration' [10] and of the creation of "a common European space" of safety and security. The fundamental question of how these emergent spaces are to be governed remains open. For several decades cross-border administrative regions have been set up, building upon economic and political experimentation at sub-national scales within

member-states, and benefiting from various funding EC mechanisms. These became locally institutionalized administrative frameworks for cross-border building, and are charged with creating a coherent trans-boundary identity by promoting networking and economic, environmental, and cultural flows. However, some have partly served as convenient administrative devices for local elites to tap into Brussels funding, with important decisions affecting the local economy taken at national and global levels. Often, local inhabitants are unaware of these initiatives, while the composition of these administrative bodies does not provide for democratic accountability.

Evaluations of cross-border cooperation, research on borders and EU assessments of the INTERREG programme reveal that border regions vary widely in terms of their origins, longevity, permeability and history of contacts, both formal and informal. They all suffer from the negative economic effects of their marginal location; nevertheless they vary regarding their potential accessibility to the core, most important regions of economic activity in the EU, as well as in terms of the immediate benefits to be accrued from transfrontier links. Internal border regions closer to the geographical and economic centre of the EU would benefit economically to a greater extent from the Single European Market abolition of internal borders [11], while the EU external borders face additional problems related to any possibility of consolidation with the deepening of EU economic integration and the widening of the economic and demographic gaps between the EU and its South Eastern European and Mediterranean neighbors [12].

2.2. The Single European Act, EU enlargement and new neighbours

The Single European Act in 1986 promised ‘a Europe Without Frontiers’ and provided a stimulus to European integration, as between 1986 and 1992, the European Commission issued almost 300 directives eliminating physical, fiscal and technical barriers to the Single European Market [13]. Significantly, national states were handed over responsibility for implementing these directives, thus allowing considerable scope for the exercise of national sovereignty and discretion in the long and difficult process of removing internal border controls [14]. Consequently, the Single Market may have rendered borders more porous by deregulation, although re-regulation is occurring carving out new roles for national states and elites within their own boundaries through the implementation of directives emanating from Brussels. The Single European Market provided the impetus for the Maastricht Treaty and its programme for closer economic and monetary union in the direction of closer political union.

Within the framework of introducing programmes and policies to improve the economic situation of the Central and East European countries, the European Union established the PHARE programme in 1989. Cross-border cooperation became a PHARE task, endowed for the period from 1994 to 1999 with about 900 million Euro, a rather low budget compared to that spent on defense issues by unilateral and multilateral actors. A direct link was established in the transition countries between border management and the setting-up of a market-oriented economy combined with democratic institution-building. Considerable attention was paid to the development of efficient border structures, mainly under the auspices of the customs and transport regulation [15]. The objective, then still traditionally phrased as “improving the effectiveness of border controls”, in 1997 switched to the more business and trade-oriented language of “effective border management”

3. Schengen-Amsterdam-Laeken

During the first years of the Schengen system operation, it became obvious that expectations were not met at the external border as the existing rules and resources turned out to be insufficient for ensuring coherent border management [16]. Nor were they met inside the territory where even the Schengen Convention’s advancing of police powers considerably lagged behind the freedom of movement granted to citizens through the Single Market. An advancement was achieved by the Amsterdam Treaty of 1997, whose Art. 62 foresaw Community competence for the crossing of external borders and rules on visas, as well as the integration of the Schengen acquis into the EU legal framework in May 1999.

At the Laeken European Council in December 2001, member States undertook to improve the management of the external EU border controls so as to combat terrorism, illegal immigration and human trafficking more effectively [17]. The Commission Communication towards integrated management of the external borders of the member states of the European Union, delivered upon the Council's request in May 2002, [18] focused on five categories of suggestions: Common corpus of legislation; Common mechanism for coordination and cooperation; Common, integrated risk analysis; Staff and inter-operational equipment; From financial burden-sharing to the European Corps of Border Guards. As regards internal measures, the emphasis was on Joint enforcement centres; Alternative checking methods; 'Security partnership' – A new generation of bilateral agreements; the European Border Agency; a Proposal for a European border code; and a Proposal for a regime of local border traffic. It was this process that led to the development of the concept of integrated border management within the EU for application at its external borders and in the zone outside its immediate borders.

The European Security Strategy (ESS), adopted in December 2003, defined key threats to the European Union of a transfrontier dimension, including terrorism, weapons of mass destruction and organized crime. Sometimes called the 'first line of defense', customs, border guards and other security personnel, if sufficiently well trained and empowered by appropriate legislation, could target, inspect, interdict and seize suspect shipments that may include proliferation-sensitive items.

The Hague Programme in 2004 furthered the discussion by adopting the assumption that there was something like 'the integrated management system for external borders' [19]. IBM, emerging as a joint concept from the PHARE, TACIS and CARDS assistance programmes, comprised a four-fold concept: (i) a comprehensive approach to border problems across; (ii) administrative; and (iii) national dividing lines under the management of (iv) dedicated professional skills. It is widely referred to by international organisations such as NATO (2003), the OSCE (2003), the Centre for the Democratic Control of Armed Forces (DCAF, 2004) [20] and in the Stability Pact [21] (2003) [22].

3.1. The Creation of 'a Common European Space'

The merging of legal systems, not even their harmonisation or approximation, was never a part of the vision of the EU founding fathers. The granting of the 'four freedoms' -- the free movement of persons, police and criminal justice agencies -- opened up the internal borders in the 1980s, while until then the European police and judicial services had been operating far apart from each other. Despite considerable progress made during the 1990s with justice and home affairs becoming the third pillar of the EU, JHA remained a step-child of European integration even at the Tampere European Council of October 1999, where instead of any full harmonisation there could just be mutual recognition of diverging criminal justice systems [23].

IBM regulations are spread across a range of legal and administrative instruments, mainly in formal legal texts such as the Treaty on the European Community or the Schengen instruments of 1985-90, and represent a multi-layered compilation of provisions. Informal arrangements, such as the 'Common Manual' on external borders adopted by the Schengen Executive Committee [24] and the 'Catalogue of Best Practices' drawn up by the Working Party on Schengen Evaluation, constitute the corpus from which much of the rest has been adopted [25].

Further elements that make the IBM mechanism work practically are found in bilateral/multilateral arrangements among individual member states or between them and third countries. This situation is characteristic of the unsecured terrain on which the European Area of Freedom, Security and Justice had to be built [26].

Most of the complications in running the EU external border are caused by the unfinished status of the EU. The border resembles a scattered line encompassing and framing a colourful patchwork of separate territories as the 'EU border' is subdivided into loosely connected national segments, each of them attended to by separate services. There is nothing such as a coherent EU territory characterised by a single legal system and protected by a single border service from coast to mountain or coast, a clearly demarcating green (land) or blue (sea) border.

The EU relies on twenty-five separate legal systems, thirteen of which are found (together with those of Norway and Iceland) within the territory of Schengen. Border management is performed by fifteen distinct national services, each exclusively responsible for one special section of it. Additionally, following the gradual abolition of internal border controls, the respective shares of the border have become increasingly unbalanced, as EU border management is not confined to the simple checking of passports and the prevention of illicit movements across the green (land) border. Staff at the border crossings have pledged to carry out their assignments taking account of the “interests of all parties” [27] and refuse entry to foreigners “representing a threat to public policy, national security or international relations of any” Schengen member. Considerable knowledge of the political/legal situation in other countries, as well as the appropriate language skills, seems thus indispensable.

EU customs – as a trade and economy-oriented administration – has always been ahead of police and border guards in streamlining European borders and the procedures related to their crossing [28].

.... The joint EU customs territory with a common external tariff was already completed by 1 July 1968, as opposed to Schengenland whose nucleus saw the light of day not before 1995. Customs administrations very early operated joint (juxtaposed) border installations (early 1960s), a Community-wide mechanism for fraud-combating (Naples Convention of 1967) and customer-oriented programmes stipulating a business-enhancement role of customs. Further, they concluded Memoranda of Understanding (MoU) with the transport community (since the 1980s) and engaged in joint vocational training of customs officials (Matthaeus programme of 1991) [29].

Following the 9/11 events, EU customs authorities started using the IBM formula predominating up to then in the police world, and to define itself primarily as inter-agency cooperation at the national and/or international level, thus establishing a seamless link with the Schengen border approach [30]. The EC welcomed, in its 2003 Communication, a simple and paperless environment for customs and trade that provides traders with equal electronic access to border procedures throughout the EU while it allows the Commission to perform a large-scale analysis of specific crime risks on the basis of the comprehensive trade data obtained.

“On top of this, a second Communication, transmitted to the Council and Parliament under the same cover, directly addresses “the role of customs in the integrated management of external borders” (ibid., p. 35), expressly underlining its complementary character in relation to the police/Schengen IBM Communication of the previous year: Both Communications are complementary and constitute the first stages in the overall strategy that the Commission is proposing for integrated and effective management of external borders, the aim being to achieve a coherent framework for joint action at EU level.” [31]

4. The development of a border management policy in the Western Balkans

At the Ohrid Conference in 2003, NATO, the EU, the OSCE and the Stability Pact for South Eastern Europe launched a new strategy for border security management in the Western Balkans, the Ohrid Border Process, further developed following the conference jointly by the states and partner organizations [32]. The general objectives specified at the conference originated from a regional conference held in Bucharest in October 2001, hosted by the Stability Pact [33]. The five countries involved in the Stabilization and Association process (SAp) (Albania, BiH, Croatia, FYROM and the then Federal Republic of Yugoslavia) were considered to be in need of assistance in order to adapt to EU standards. The objectives agreed on were clearly demarcated borders; good cooperation with neighbouring states; non-military border guarding structures; a sound and legal framework based on EU standards; well-trained, -equipped and -managed services; and smooth bilateral and multilateral cooperation in border guarding.

Regarding the initial development of national instruments, it was decided that a definition of national policies and strategies on IBM would be established. The Commission-funded CARDS IBM programme, which deals with the interrelated issues of trade facilitation, border control and the

development of border regions in the Western Balkans, was given responsibility for this task. National coordinating structures and procedures were to be implemented, and where required identification of command and control arrangements through appropriate channels. On a regional basis, internal, bilateral and multilateral mechanisms and procedures for the exchange of information on border and trafficking issues were introduced. A training programme for military personnel directly involved in border control management and anti-trafficking activities was also initiated [34].

Other international non-State actors are present in the IBM world of South East Europe too, for example the International Organisation of Migration (IOM), mainly working on streamlining processes of passports and documentation. In an article in the New York Times about the use, or lack of use, of Interpol data on stolen passports, the case of Mr. Milorad Ulemek, a suspect in the assassination of Serbian Prime Minister Djindic, was raised. The article pointed out that the suspect traveled frequently in 2003 and 2004 through six countries on stolen passports, some of which were recorded in the Interpol database. The passports were stamped 26 times, with no detection of the stolen documents he was using [35].

Another intermediary, the International Centre for Migration Policy Development (ICMPD) [36], focused on regional border guard cooperation in a long-term EU/Schengen perspective. Since the first Ministerial conference of 1993, the Budapest Process has dealt extensively with issues of relevance to border guarding, especially following the adoption of the Ministerial recommendations in Prague in 1997 and the follow-up activities initiated in the years thereafter. Particular emphasis lies on the exchange of information and sharing of best practices in the area of border control for the prevention of irregular movements in general and trafficking and smuggling in particular. The activities related to border guarding increased in 1998, when the Working Group on South East Europe was created within the Budapest Process. International examination teams under this Working Group were sent out to examine land- and sea-borders and airports, in Albania, Bosnia and Herzegovina, Bulgaria, Croatia and Romania. This Working Group in 2001 became the Border Guard Task Force for South East Europe, established within the Stability Pact for South-Eastern Europe. Table III of the Stability Pact, it was decided that the activities of the working group would be subsumed by the Ohrid Border Security Process and to some extent also by MARRI, the Migration, Asylum, Return Regional Initiative of the Stability Pact [37].

5. How does IBM and CBC travel to South East Europe?

The European Commission has become the primary sponsor of a range of transnational and cross-border institutional initiatives from the Danube to the Mediterranean;

working through the principles of ‘partnership’ and ‘subsidiarity’, transnational spatial planning, operating within an inter-governmentalist framework, is meant to usher in a Europe conceived as a purported ‘multi-level polity’, guided by a multiplicity of actors including the European Commission, national governments, urban and regional administrators, public/private partnerships, universities and elements of civil society [38].

Most of the targeted border regions are situated where the EU borders with the Western Balkans, like the Carpathian region, and/or in areas that are perceived as contested zones of conflict, such as the Southern Adriatic, composed by border areas of Montenegro, Croatia and Bosnia and Herzegovina. Three border regimes are suggested by Kramsch and Hoper, viewed as tendencies capturing modes of governance which might overlap, producing hybrids, namely the absent (non)border, the marsh and the postcolonial limes [39]. Attempting a comparison of research findings on EU cross-border institutionalisation experiments and the SEE borders, shows that any clear cut categorisation of border regions is indeed difficult and the demarcating lines between types are blurred.

The border type category of ‘marsh’, a term denoting a fief in between the king’s territories and *terrae incognitae*, a buffer zone, partially fits the picture. Within this logic of cross-border governance, the Accession countries of Central and Eastern Europe are positioned as newly-minted buffer zones, reminiscent of the older term ‘Mitteleuropa’. A closer examination of the difficulties to regulate

migration in the Balkans is indicative. All countries involved have a long experience with migration, while in the cross-border regions deep histories of very different experiences of statehood, as well as very old and complex in-, out-, and through-migrations are played out. The types of mobility are not exclusively national and local, as this region is also a transit zone, offering a market and services to those migrating. The local demand for easier seasonal passes for the cross border areas' inhabitants vis a vis the strict regime of Schengen visas, the calls to create 'smart borders, smart visas', as well as the recent international attention to developed asylum legislations correspond to the newly perceived security threats coming from countries beyond the existing EU border, transboundary criminal networks and illegal migrant flows.

In practice there are substantial restrictions on the freedom of travel for example of citizens of the Western Balkan countries when they attempt to travel into the EU. Because of high visa fees, only one per cent, for example, of these countries' students have had an opportunity to visit an EU country. In addition citizens, awaiting their visas in the diplomatic missions of EU states, have often been mishandled and humiliated. A comprehensive facilitation of the allocation of visas condition was for example that the WB countries' governments should take wide-ranging measures for the defense against migration. These measures would include the countries' fulfillment of the EU's expectations of a more consistent application of the so-called readmission agreement and the conclusion of these contracts with all EU states. Germany had already concluded a readmission agreement with FYR Macedonia in 2002. It has been in effect since 1 May 2004 and provides for Skopje's unconditional acceptance of the deportation to FYR Macedonia of anyone without a valid residence permit. For relaxation of visa requirements, the EU requires the government to make the means available for the fortification of its borders and issue new passports, all requiring the involvement of foreign companies.

The difficulties faced by all actors, national, international and regional, in addressing visa facilitation within the region are best exemplified in the following declaration of the Migration, Asylum and Refugee Regional Initiative (MARRI) at the Regional forum held in Tirana on 5 April 2005:

... 1. Taking into account the objectives of the EU-Western Balkans Summit Thessaloniki Declaration, we shall continue the discussion on further cooperation among our countries and the EU, in order to facilitate our respective visa regimes, having in mind national policies and the compatibility with the relevant EU policy and requirements.

2. We shall enter into regional and bilateral discussions in order to enhance consular cooperation building upon the results of the EU CARDS Program. This process shall be focused on exchange of information, including warning and mutual alerting on suspected illegal transit flows and consular cooperation, according to best practices offered by the EU. A regional working group to facilitate these discussions will be established, consisting of representatives of the responsible ministries, and will convene regularly at the MARRI Regional Centre.
... [40].

Thus efforts to establish cross-border institutional initiatives clash with the exclusionary logic of Schengen. Kennard and Virtanen, looking into cross-border initiatives encompassing spaces of the Ukraine, Poland, Slovakia, Belarus, Finland, Russia, Hungary, Romania, Lithuania, Germany and Sweden, examine how the EU speaks with a 'forked tongue': the Schengen regime and the Committee of the regions – the latter under whose auspices cross-border regions are developed – are presented as:

“ initiatives designed to overcome ‘bad’ national borders that inhibit integration and cohesion. These include the development of such cross-border ventures as tourism, university programs, and fairs to encourage regional trade” (p. 13 [41]).

The guarding of the borders against 'aliens' becomes a priority, thus the intensified policing of the EU borders-to-be is directed towards the prevention of the border crossings of non-European individuals, both at the Eastern and the Balkan 'marshes'. Kennard and Virtanen also examine the historical as well as the geographical complexity of the cross-border regions and the continued influence of 'border-memories' which may not necessarily be the same for the successive waves of occupants living in the region. In the WB region as well there are no pure borders and no pure identities, as

borders crossed over the heads of people several times. The question is raised whether it is possible to use cross-border regions as sites or tools to regenerate peripheral regions and thus to decrease economic disparities between Europe's different parts. Once the logic of a link between an identity and a territory is broken as an instrument in negotiating conflicting interests, whether a common identity is indeed a necessary condition for acting in common is problematised, ultimately posing the question as to the necessity of a 'common European identity'.

Yet any kind of normative approach seems risky. South East Europe is an extremely rich space of emerging transnational governmentality, of international security discourse and of EU influence. The processes of policy transfer and negotiation and the alleged emergence of new governance structures is being discussed as Europeanisation, namely

the emergence and development at the European level of distinct structures of governance, that is, of political, legal and social institutions associated with political problem solving that formalizes interactions among the actors, and of policy networks specializing in the creation of authoritative European rules. [emphasis original] (p. 3, [42]).

Europeanisation has been employed to refer to examples where distinct European forms of organisation and governance have been exported outside Europe's territorial boundaries (p. 14, [43]).

In the context of globalisation and Europeanisation the analytical framework for understanding national policy change has changed to incorporate the influence of external actors [44]. In particular the understanding of post-communist policy making has been advanced by studies focused on the reforms within the first wave of countries to join Europe [45]. In this processes a whole range of actors is involved, while concerns and debates about transnational consultancy and in particular of the role of policy consultants, experts and 'brokers' are emerging. Who are these intermediaries and what do they do? Are they a parasitic class sitting upon a big market called South East Europe? Or are they the champion of change? Entrepreneurs who engineer new visions? Examining the opinion of two European policemen holding posts with the EU Proxima [46] Police force in the Ohrid area during 2004 provides useful insights of the possibilities:

We do not understand why the name Macedonia is such a huge issue for either side; it is not in Alsace or in Switzerland. Sometimes we are aware that Albanians cross into FYR Macedonia, issuing their visa in Stenje on the spot, according to the Albania-Macedonia bilateral agreement, with the intention of crossing over the mountains into Greece, now that the Greek-Albanian border is well guarded. But it is difficult to get the officers from the two respective borders, Niki and Medzitlija, to pick up the phone and talk to each other. That is a great impediment to illegal migration, trafficking and smuggling interception. It is our dream to organise joint seminars and workshops for border guards and border police from all sides [47].

At the same time, an Albanian policeman at the Krystalopiyi cross point sounds quite satisfied with developments while checking passports of over twenty persons waiting crammed in front of his booth:

We are becoming computerised, allowing for better interception of criminals. So many travel by busses, lots of private companies operating them as well, and many come by private cars. We caught many criminals who came from Greece and Italy to watch the football game between Albania and Greece in Tirana, and who were not aware that now we have computer systems and we are connected with the Europol. We got over twenty of them [48].

6. Transcending borders?

An overarching critique of mainstream (Western European) approaches to these themes is provided by József Böröcz in his discussion of 'Empire' and 'Coloniality' in the 'Eastern Enlargement' of the EU. He points out how narrow has been the focus of the question 'What is the EU today?', as a completely internal matter such that any discussion of 'enlargement' is simply added on, almost as an afterthought. What he terms "the traffic between the internal and the external" of the EU, or the

constructed dichotomy between the ‘falsely inclusive’ and the ‘falsely exclusive’, when viewed through the lens of empire, coloniality and orientalizing discourses, shows how the EU is “a constitutive focus and center of dependence for important social, cultural, economic, and political processes at places outside of Western Europe” (p.6 [49]).

Böröcz’s tracing of the ‘shared-but-different’ histories of empire and associated identities through coloniality yields empirical expectations to be further investigated

that the specific histories of colonialism and empire, with their deeply coded and set patterns of inequality, hierarchy, exclusion and power – and especially their techniques pertaining to the projection of that power to the outside world – are reflected in a deep and systematic form in the socio-cultural pattern of the governmentality of the EU (p.14, [50]).

The importance of understanding this in more plural terms, as governmentalities, appears important here, as does Clarke’s rejoinder that people always live with/in and against particular governmentalities which always enter ‘national-popular formations’ only in and through alliances, ‘assemblages of political discourses’ which inevitably change, shape, and produce ‘hybrids, paradoxes, tensions and incompatibilities’ rather than “coherent implementations of a unified discourse and plan” (p. 94, [51]).

The language used when referring to IBM does not necessarily always match the one applied to security considerations in most texts of international organisations. One of the intermediaries in South East Europe, the EastWest [52] Institute’s Global Security Programme, in co-operation with its Regional and Transfrontier Co-operation Programme, convened a policy meeting of 48 participants from approximately twenty countries, which produced the following document quotations:

“Integrated border management is an element of the EU’s strategy for controlling borders in the Balkans whilst accounting for the significance of border regions, cross-border cooperation and conflict prevention. It focuses not simply on borders per se but also on the cross border flow of persons and goods. The ambition is to provide security and freedom by getting past the notion that security emerges from tight border control. Integrated Border Management relies on constant evaluation and improvement to ensure the greatest level of quality control.

Elements of an EU Integrated Border Management Strategy to achieve Schengen capability within the next 10-15 years are: 1) reinforcement of border security through EU-standardised border surveillance and control (an example of the new tamper-proof visa is shown at right); 2) facilitation of border crossing traffic through EU-standardised control procedures in border checkpoints, 3) enhancement of alien issues through EU-standardised aliens’ policy, 4) improvement of inter-agency co-operation through a co-operation network both internally with all governmental agencies and economic partners involved in border administration and externally with corresponding agencies of neighbouring countries and 5) the creation of a new legal basis through legislation in compliance with the EU *Acquis Communautaire* [53].

Europeanisation, policy formulation and implementation though is by no means a simple singular linear process from planning to outcomes. Instead it is more appropriate to conceptualise the implementation of policy as the outcome of a contest for resources, influence and meaning between different actors. In order to understand the process of policy implementation and the links between these different sites it was necessary to ‘study up’ to investigate “processes whereby power and responsibility are exercised”. The definition and description of the context completed, the text carries on with a set of questions related to the issue of agency and agencification:

... Integrated Border Management is an evolving concept and could serve as an increasingly important tool in preventative diplomacy. Which international organisations are best suited to developing this powerful tool and bringing it to new regions? The EU? NATO? The Council of Europe? The OSCE? Each organisation has some of the qualifications and competencies that this would require. These questions raise more questions of a basic nature as one attempts to

determine the best sort of border regime for a given place as well as *the organisation best suited to promoting it*.

- What and where are the borders? What is on the other side of the border? A suitable partner?
- How long will these borders endure? Will there be integration in Schengen and, if so, when?
- Who will guard the borders? Will the guards be military, police or paramilitary?
- What kind of international assistance is available? Where is the most valuable expertise and what organisation should we invest in? Which institution should take the lead? The EU, NATO or the OSCE? On this question there was a lack of agreement for each institution has different competencies and different limitations.
- What motivations should countries in the Wider Europe have? What interests are addressed?
- How will it be possible to develop a real implementation strategy? (emphasis in the original) [54].

Wedel addresses the importance in these encounters of multiplex networks where players know each other, and interact, in a variety of capacities, with multiple identities (which she terms ‘transidentities’), and in a variety of roles. Her tale is one of shifting and multiple agency, promoted in part by what she terms ‘flex organisations’, which have a ‘chameleon-like, multipurpose character’, with actors within them ‘able to play the boundaries’ between national and international; public and private; formal and informal; market and bureaucratic; state and non-state; even legal and illegal (pp. 149-174, [55]).

Ferguson and Gupta have argued that states are increasingly challenged in terms of their ability, through routine bureaucratic practices which produce forms of verticality and encompassment, to present themselves as ‘legitimate’, even ‘natural’ authorities. Seeking to study the relationship between states, space and scale as part of a wider ethnography of neo-liberal governmentalities, Ferguson and Gupta suggest that both ‘supranational’ and ‘grassroots’ actors are engaged in similar practices of verticality and encompassment. Hence, they are critical of a perspective which treats the ‘global’ “... as if it were simply a superordinate scalar level that encompasses nation-states just as nation states were conceptualized to encompass regions, towns, and villages” (p.990, [56]).

They continue:

... it is necessary to treat state and nonstate governmentality within a common frame, without making unwarranted assumptions about their spatial reach, vertical height, or relation to the local ... For the central effect of the new forms of transnational governmentality is not so much to make states weak (or strong), as to reconfigure states’ abilities to spatialize their authority and to state their claims to superior generality and universality (pp. 994 and 996, [57]).

The INGO EastWest Institute text concludes with a paragraph referring to historical processes and borders almost as to stages in a linear, psychologically determined, almost metaphysical and teleological order, in which the same actors, states, organizations play the leading roles:

“From an historical perspective, it is important to understand the *three “moments” of borders*, namely border creation, border maintenance and border transcendence. *Border creation* is frequently a bloody phase in which the instruments include coercion, dynastic plotting and warfare. It is rarely a democratic moment. Moreover, with the relative decline of state sovereignty and concomitant rise of transnational norms and organisations, border creation is a less absolute concept than it was a century ago. *Border maintenance* is tremendously complex and relies on the work of numerous kinds of national and international organisations. This implicates defence establishments, border control agencies and multilateral organisations. *Border transcendence* requires the most creativity and offers the greatest hope for the development of a greater sphere of peace and prosperity. It can be accomplished by international / bilateral or transnational efforts. This work

rests on different and potentially contradictory conceptions of borders: as barriers, as bridges or filters, as resources (especially where an asymmetry of wealth across borders permits arbitrage), and as symbols. To put it differently, viewing borders as frontiers encourages efforts to transcend them whilst viewing them as symbols of national sovereignty implies a need to bolster them. In the end, *building trust across borders and co-operation in border regions* will generate the confidence necessary to transcend borders.” (emphasis in the original) [58].

7. Conclusion

While conducting fieldwork along the Balkan euroregions under construction, nation-state territoriality as a container-limit for politics, is problematised. The question arises as to how the contradictions inherent to the ‘governmentality’ of EU enlargement apply in the South East European borders case, particularly as the imperatives of market expansion collide with the security goals of Schengen. The discourse about the reasons for controlling these transfrontier areas resonates with the more general EU discourse on borders, referring to problems of maintaining territorial security and the ability to control flows in the context of threats stemming from transnational crime, trafficking, and terrorism. The Commission is increasingly concerned with the organization of non-EU space as made evident in the recent European Neighbourhood Policy. There is a sort of Europeanisation of space, in which EU governance is concerned with the construction and management of European spaces and networks, distinct from the territorial places and spaces characteristic of the nation-state. EU borders’ policies are in flux, caught between forging links across the external border and policing it. Signs of internal challenges to the bordered sovereignty of member states co-exist with the dynamic nature of the EU’s external borders.

Acknowledgments

I want to thank my supervisors Bob Deacon and Paul Stubbs for their continuous support and encouragement throughout my PhD research, of which this article is a small part.

- Bob Deacon, University of Sheffield, Dept. of Sociological Studies, Elmfield, Northumberland Road, Sheffield S10 2TU, B.Deacon@sheffield.ac.uk
- Dr Paul STUBBS, Senior Research Fellow, The Institute of Economics, Zagreb, pstubbs@eizg.hr

References

- 1 Alkan, N. “Borders in Europe” in N. Alkan (ed.) European Integration and South Eastern Europe 2002, Bonn: Center for European Integration Studies.
- 2 Duffield, M. (2001) Global Governance and the New Wars: the merging of development and security. London: Zed books.
- 3 <http://www.seccenter.org/>
- 4 Kramsch, O. and Hoper, B. (2004) eds. Cross-Border Governance in the European Union. London: Routledge.
- 5 Dunford, M. 1994. Winners and losers: the new map of economic inequality in the European Union. European Urban and Regional Studies 1 (2): 95-114.
- 6 Hobbing P., Integrated Border Management at the EU Level, CEPS Working Document No.227, August 2005.
- 7 Strassoldo, R. (ed.) 1973. Boundaries and regions. Trieste: Edizioni.
- 8 Strassoldo, ibid; and Strassoldo, R. and G. Delli Zotti (eds) 1982. Cooperation and conflict in border areas. Milan: Franco Angeli Editore, Anderson, B. 1992. The new world disorder. New Left Review 193, 3-13.
- 9 Development of the Community’s territory. Brussels: European Commission, p. 169.

- 10 Kramsch O. (2003) The Para-Site of Governance: Transborder Regionalism in the Euregios, WPTC-01-03, Nijmegen Centre for Border Research.
- 11 Nijkamp, P. 1993. Border regions and infrastructure networks in the European integration process. *Environment and Planning C: Government and Policy* 11: 431-446; Jones, P.N. and Wild, T. 1994. Opening the frontier: recent spatial impacts in the former inner-German border zone. *Regional Studies* 28 (3): 259-273.
- 12 Therborn, G. 1995. *European modernity and beyond: the trajectory of European societies 1945-2000*. London: Sage, p. 244.
- 13 Williams, A.M. 1994. *The European Community*. London: Blackwell, p. 110-115.
- 14 Butt Ph, A. 1991. European border controls: who needs them? *Public Policy and Administration* 6 (2): 35-54; Lodge, J. 1992. Internal security and judicial cooperation beyond Maastricht. *Terrorism and Political Violence* 4 (3): 1-29.
- 15 Van Kuik, 2001 and European Commission, 1998 (Consortium Eurocustoms, created in 1991 by the 12 EU customs administrations).
- 16 European Commission, 2002, Communication, Towards integrated management of the external borders of the Member States of the European Union, COM(2002) 233 final, Brussels, 7.5.2002, paras. 12-14.
- 17 European Council, 2001, Presidency Conclusions of the Laeken European Council, 14-15 December 2001, SN 300/1/01, REV I, Conclusion No. 42.
- 18 European Commission, 2002, Communication, Towards integrated management of the external borders of the Member States of the European Union, COM(2002) 233 final, Brussels, 7.5.2002.
- 19 European Council, 2004, The Hague Programme: Strengthening Freedom, Security and Justice in the European Union, Council Doc. 16054/04 of 13 December, item 1.7.
- 20 The Geneva-based DCAF dedicates, as part of the security-sector reform in the Western Balkans, considerable resources to the restructuring of border services in the countries concerned. The 'EU requirements' are explicitly referred to as a benchmark for all these reforms.
- 21 The Stability Pact was initiated by the European Union in May 1999. It was subsequently adopted at an international conference held in Cologne on 10 June 1999 and placed under the auspices of the OSCE. It is designed to contribute to lasting peace, prosperity and stability in South Eastern Europe through coherent and coordinated action, by bringing together the countries of the region, other interested countries and organisations with capabilities to contribute. It establishes specific mechanisms to coordinate their joint efforts, www.stabilitypact.org.
- 22 Peter Hobbing, 2005, *ibid*.
- 23 European Council (1999), Presidency Conclusions of the Tampere European Council, 15-16 October 1999, SN 200/99.
- 24 Council of the European Union, 2002a, Common Manual, OJ C 313 of 16.12.2002, p. 97.
- 25 Catalogue of Best Practices on External Borders Control, Removal and Readmission: Recommendations and Best Practices of February 2002, (the text is "explanatory and having no legally binding status"), <http://ue.eu.int/uedocs/cmsUpload/catalogue20EN.pdf>
- 26 Hobbing, 2005, *ibid*.
- 27 Art. 6, Schengen, 1990.
- 28 Hobbing, P. (2003), Management of the External Borders: Enlargement and The European Guard Issue, DCAF Conference Managing International and Inter-Agency Cooperation at the Border in Geneva 1-15 March 2003 (http://www.dcaf.ch/news/Border%20Mgt_031303/Hobbing.pdf).
- 29 Council Decision of 20 June 1991, Official Journal L 187, 13.07.1991, p. 41.
- 30 WCO, 2004, Resolution of the Customs Co-operation Council on global security and facilitation measures concerning the international trade supply chain (June 2004).
- 31 Hobbing, 2005, *ibid*.
- 32 'The way forward', presented at the Ohrid Regional Conference on Border Security and Management, 22-23 May 2003.
- 33 The National and Regional Management and Development of Border Control, Regional Conference, Bucharest, 25-26 Oct. 2001.
- 34 Ian Anthony, Aline Dewaele, Rory Keane and Anna Wetter, Strengthening WMD-related border security management assistance, September 2005, Stockholm International Peace Research Institute.

- 35 www.iom.org, Charles Harns, Head, Technical Cooperation International Organization for Migration – IOM Strengthening Border Management and Security: Developing a more Strategic and Coordinated Approach to Technical Assistance, JOINT OSCE/UNODC INTERNATIONAL TECHNICAL EXPERTS CONFERENCE ON BORDER MANAGEMENT AND SECURITY, Vienna: 7 – 8 September 2004.
- 36 www.icmpd.org.
- 37 ICMPD Activities in the field of Integrated Border Management, July 2005, www.icmpd.org.
- 38 Kramsch and Hoper eds, 2004, *ibid*.
- 39 Kramsch and Hoper eds, 2004, *ibid*.
- 40 MARRI Regional Forum Declaration, 2005, <http://www.emins.org/dokumenti/050504-marri.pdf>.
- 41 Kramsch and Hoper, 2004, 13.
- 42 Risse, T., Cowles, M. G. and Caporaso, J. “Europeanization and Domestic Change: Introduction” in M. G. Cowles, J. Caporaso and T. Risse (eds.) *Transforming Europe: Europeanization and Domestic Change 2001*, NY: Cornell University Press.
- 43 Buller, J. and Gamble A. *Conceptualising Europeanisation. Public Policy and Administration* 2002, 17(2)
- 44 O’Brien R et al 2000, *Contesting Global Governance: Multilateral Economic Institutions and Global Social Movements*, Cambridge: Cambridge University Press and Boas M and McNeill D 2004, *Global Institutions and Development: Framing the world?*, London: Routledge
- 45 Deacon et al 1997, *Global Social Policy*, London: Sage
- 46 The European Union established an EU Police Mission in the former Yugoslav Republic of Macedonia (fYROM), in line with the objectives of the Ohrid Framework Agreement of 2001 and in close partnership with the country's authorities. The Mission, code-named EUPOL PROXIMA, was launched on 15 December 2003 (Council Joint Action 2003/681/CFSP). EU police experts were monitoring, mentoring and advising the country's police thus helping to fight organised crime as well as promoting European policing standards. PROXIMA was part of the European Union's overall commitment in assisting the efforts of the Government of fYROM to move closer towards EU integration. This operation was completed on 14 December 2005, http://www.consilium.europa.eu/cms3_fo/showPage.asp?id=584&lang=EN&mode=g.
- 47 Interview, November 2004.
- 48 Interview, September 2004.
- 49 Böröcz, J. (2001) ‘Introduction: Empire and coloniality in the ‘Eastern Enlargement’ of the European Union. in Böröcz, J. and Kovács, M. (eds.) *Empire’s New Clothes: unveiling EU enlargement*. Telford: Central Europe Review.
- 50 Böröcz, J. (2001), *ibid*.
- 51 Clarke, J. (2004) *Changing Welfare, Changing States: new directions in social policy*. London: Sage.
- 52 “EWI is an independent, not-for-profit, European-American institution working to address the most dangerous fault lines of the 21st Century and to help build fair, prosperous and peaceful civil societies in those areas. Since 1981, we operate long-term projects that create trust and understanding and seek to reduce tensions from Eurasia to the trans-Atlantic region using our unique network of private and public sector leaders in more than 40 nations. Our initial mission was devoted to reducing the hostilities of the Cold War”, <http://www.iewe.org/>
- 53 EWI Conference Report: BUILDING AN AREA OF SECURITY AND FREEDOM IN THE WIDER EUROPE THROUGH INTEGRATED BORDER MANAGEMENT, March 19, 2004, The Centre, Brussels, www.ewi.info.
- 54 EWI Conference Report, *ibid*.
- 55 Wedel, J. (2004) “‘Studying Through’ a Globalizing World: building method through ethnographies”. In Gould, J. and Secher Marcussen H. (eds.) *Ethnographies of Aid: exploring development texts and encounters*. Roskilde University IDS Occasional Paper 24.
- 56 Ferguson, J and Gupta, A. (2002) ‘Spatializing States: towards an ethnography of neo-liberal governmentality’ *x.mn z.xcn*, in *American Ethnologist* 29 (4).
- 57 Ferguson, J and Gupta, A. (2002), *ibid*.
- 58 EWI Conference Report, *ibid*.

Talking about Migration in Central Northern Greece: The Employment of the Construct of Greek Emigrants

Maria Xenitidou¹, Professor Paul White³, Dr. Sean Homer²

¹SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, mxenitidou@seerc.info

²City College, Dept. of Business and Economics, 13 Tsimiski Str, 54624 Thessaloniki, Greece, s.homer@city.academic.gr

³The University of Sheffield, Dept. of Geography, Western Bank, Sheffield, S10 2TN, UK, P.White@sheffield.ac.uk

The geopolitical changes of 1989 saw the development of multiple forces, of which migration is one. Greece experienced a shift from a traditionally sender country to a main destination country for immigrants from Central and Eastern Europe and the former Soviet Union. Employing a discursive social constructionist approach, this paper presents the functions of the employment of the construct of Greek emigrants abroad in the context of Greek people from Central Northern Greece talking about migration. Namely, embarking on the premise that identities are discursive actions, which are socially constructed in interactional contexts, this paper explores the subject positions interacting individuals negotiate when they mobilize the construct of Greek emigrants in the context of talking about migration and the function of this mobilization.

Keywords

Discursive social constructionism, Greece, migration, subject positions.

1. Introduction

The period following the 1990's has been characterized for calling into question previous understandings of social, economic and political identity in Europe. Greek accounts of national identity seem to be informed by a number of recent forces, of which migration from the Balkans is one. The seeming geopolitical stability in the second half of the 20th century and the tightening of migration regimes in northern Europe saw a relative halt in mass immigration flows, which was interrupted since the 1990s. Greece became a destination country for immigrants from Central and Eastern Europe and the former Soviet Union. According to the 2001 census, 7% of the legal(ized) population of Greece are 'foreigners'/immigrants [1], while it is estimated that 2-3% appears not to have registered, amounting in total to 10% of the (adjusted) population. Half of the registered immigrants come from Albania, followed by Bulgaria, Romania, Georgia, Pakistan, Ukraine, Poland, Russia. Twelve percent of this population has settled in Central Macedonia (Central Northern Greece) and 47.5% in Attica. The percentage contribution of immigrants to the total population is 17% in the Municipality of Athens and 7% in the Municipality of Thessaloniki [2]. Nevertheless it should be stressed that, according to the same survey, the percentage contribution of immigrants from the Balkans to the total population of Central Macedonia is higher than in Attica, which is the main reason for focusing on Central Northern Greece in this research work.

The ongoing PhD research aims to explore the ways in which elements of Greek identity have taken on board the presence of 'new' migrant populations from the Balkans in Central Northern Greece. Employing a discursive social constructionist perspective, the main assumption of this research is that

national identity is a social construct negotiated, produced and reproduced in interaction, as a form of life by appealing to spatiotemporally available discourses.

The focus of this paper in particular is to present the employment and function of the construct of Greek emigrants abroad in the context of Greek people in Central Northern Greece talking about migration. The paper develops the philosophical underpinnings and methodology of this study and the presentation of the strategy of the employment of the construct of Greek emigrants abroad in the group discussions of Greek people in Central Northern Greece talking about migration.

2. Research Approaches

2.1 Social Constructionism, Discourse Analysis and Identity – Theoretical Concepts

Using a discursive social constructionist approach, this research combines concepts and tools primarily from *discourse and conversation analysis*, both from the *relativist* and the more *critical* approaches. Consequently, I focus on *patterns within language* and *patterns of activity*, on the *process* and *content* of talk-in-interaction, on the *micro* and *macro* levels of analysis respectively. The combination of social constructionism and discursive psychology is guided by the perspective that psychological phenomena e.g. memory, identity, become something people *do* (*discursive actions*) rather than something people *have* (*cognitive processes*).

Justification, rationalization, categorization, attribution, naming and blaming are discursive practices used by people in particular contexts to achieve social and interpersonal objectives (p.91, [3]).

In line with this, I draw upon *national identity as a social construct negotiated in (interactional) contexts* from which the person cannot be separated [4]. The person is immersed in the social world and draws upon discourses, which are historically, contextually, culturally and spatiotemporally available [5]. Therefore, knowledge is situated (context specific) and relative and any account is local and value-laden [6].

Secondly, discourses have a functional character as social actors draw upon them in negotiating and performing their identity. However, I extend this definition and draw upon the Foucaultian notion of discourses as being *productive* and *constitutive*, as having force and regulating practice [7]. Discourses make claims to the ‘truth’ but are not all equally powerful. Dominant discourses are *normalized* and constitute *common sense ideologies and forms of life*. Discourses on national identity or stereotypes, for instance, establish the norm through a process of contradiction, comparison and differentiation with counter-discourses. The homogeneity or shared social understanding produced is daily lived in the world of nation-states, which resembles Bourdieu’s concept of habitus, internalized through national socialization [8]. The process of *normalization/naturalization* is a central assumption of this research as regards creating awareness and sustaining identities. Namely and employing Billig’s argument, nations not only have to be imagined and narrated but they also need to be flagged in everyday life. Nationhood is asserted through universal codes such as particular national labels, flags, identities, anthems, histories [9].

In this way, ‘we’ imagine ‘ourselves’ and ‘foreigners’ to be equally ruled by the sociology of nationhood (p. 3, [9]).

2.2 Research Tools

The tools selected for this research draw on a combination of traditions in discursive social constructionism. This paper takes on board that individuals make use of *rhetorical strategies* in order to gain legitimacy. Billig argues that all discourse is rhetorical,

it is argumentative and seeks to persuade; as such the activities of criticism and justification are central to rhetorical discourse (p. 214, [10]).

The rhetorical strategies (such as Disclaimers, Humor, Mitigation strategies and Hedges, Extreme Case Formulations, Appeal to Personal Experience, Comparison, Impersonal Structures and Direct Speech/Active Voicing) used by participants in focus groups are studied in terms of their type, organization and function in undermining alternatives and persuading [11] [12] (see Appendix 1 for a brief definition of the main rhetorical strategies). This is also expected to constitute a resource as regards to the embeddedness of rhetorical strategies in (particular) ideological discourses of national identity as well as to the constant negotiation of *subject positions*.

The negotiation of positions in order to produce accounts is termed ‘positioning’ [13]. In *interactive positioning* individuals’ utterances position other interacting participants, while in *reflexive positioning* individuals position themselves. This is neither necessarily intentional, nor constrain-free, as subject positions are made available in interaction according to spatiotemporally available discourses. In other words, the structure of the language that we are born and socialised into determines the way we perceive things and live them out in everyday interactions [14]. However, Davies and Harré argue that individuals have a room for manoeuvre and choice, through awareness and uncovering the meanings within discourses [13].

In addition, linked to the concept of positioning and managing subject positions in talk-in interaction are the concepts of footing and face. Footing as developed by Goffman refers to alignment in relation to what is occurring in interaction [15]. In other words, footing refers to how one relates to what he/she says. Goffman expanded the basic ‘dyad’ of interaction – speaker-hearer – to the speaker as i) *Animator*: the person who makes the sounds, ii) *Author*: the person who selected the words and the phrasing and the iii) *Principal*: the person who is responsible for the opinions expressed¹. Participants’ footings shift over the course of interaction. Face as defined by Goffman is

the positive social value which a person effectively claims for himself by the line others assume he has taken during a particular contact. Face is an image of self delineated in terms of approved social attributes (p. 5, [16]).

In other words, a person maintains face when the his/her utterances are consistent with the verbal or non-verbal acts which create his/her image, may lose face in case of utterances inconsistent with his/her image or may be out of face when not prepared to be consistent with the line expected of him/her.

The tools that the ongoing PhD takes on board beyond the bounds of this paper are briefly presented. Firstly, apart from the appeal to broad discourses, I also focus on the process of talk-in-interaction in focus groups to explore the range of *interpretative repertoires* utilized by participants. These are

the building blocks of conversation, a range of linguistic resources that can be drawn upon and utilized in the course of everyday social interaction (p.198, [17]).

Secondly, words and utterances [13] acquire meaning within contexts or, to use Billig’s term, carry an ideological history, in the sense that they are attached to broader discourses. *Ideological discourses* contain tensions or contrary themes and are *dilemmatic*, which according to Billig provides for the premises for common sense to evolve in Western cultures, through discussion or counter-positioning [17]. Thirdly, I also draw on Garfinkel’s argument that actions are meaningful through shared understandings, and that social action requires an analysis of how social actors use shared common-sense understandings and shared methods of reasoning in interaction [18]. Joint construction - *jointly produced storylines* and *joint remembering/forgetting* - becomes practical activity, to which social actors consent, position themselves and are positioned by others and by ideological discourses.

¹ The hearer is classified into the i) primary addressee: a ratified hearer and ii) the overhearer.

3. Methodology

3.1 Sampling Parameters

In terms of sampling the target population of Greek people residing in Central Northern Greece, the ongoing research focuses on two main parameters; namely *locality*, as defined by the distinction between *urban/rural* spaces, and *age*. The *urban/rural* distinction builds on and is expected to highlight issues occurring from both societal differences per se and from the differences in the percentage concentration of migrants, while the age groups identified are: i. *18-21 (as growing alongside 'new' migration)* and ii. *35-45 (as recipients of 'new' migration)*. The urban/rural distinction is set for the purposes of this research in the Prefecture of Central Macedonia (Central Northern Greece) on the grounds that the percentage concentration of immigrants to the total population in the area is significantly high. The municipalities selected are: Thessaloniki (6,5% of foreigner² concentration), Chalkidiki (8,5% of foreigner concentration) and Serres (2,4% of foreigner concentration).

3.2 Data Collection

Focus groups have been selected in order to obtain a variety of perspectives about the topic through argumentation, positioning and counter-positioning. The benefits of focus groups include gaining insights into people's shared views and perceptions of everyday life and the ways in which participants interact with each other in a group situation (positioning). Discussion in the focus groups has been introduced by a paragraph with the intention to position participants. The role of the researcher was restricted to clarifications in order to retain natural settings. Overall, 8 focus group sessions were held with 38 participants. Each of these sessions included 4-5 participants and was controlled for gender (20 men, 18 women), education and employment status.

3.3 Analysis

Transcriptions are discourse analysed focusing on the rhetorical strategies used, their functions and dynamics within the context of discussion. This paper specifically presents the strategy of using the construct of Greek emigrants abroad, as employed by participants in 4 focus groups – 3 in the urban area of Thessaloniki and 1 in the rural area of Chalkidiki, of which 2 aged 18-21 and 2 aged 35-45 - in the context of talking about migration. Namely, focus groups 1 (FG 1) and 2 (FG 2) were held with participants from Thessaloniki aged 18-21, focus group 4 (FG 4) was held with participants from Thessaloniki aged 35-45 and focus group 7 (FG 7) was held with participants from Chalkidiki aged 35-45 (see allocation of extracts below). It should be noted that extracts were selected as representative of this strategy used in 4 out of the 8 focus group sessions held.

4. The employment of the construct of Greek emigrants abroad

The construct of Greek emigrants abroad – primarily to Germany, the USA and Australia – is extensively used in the transcripts either in the form of analogy or comparison/contrast to (Albanian) migrants and seems to serve specific functions. It should be noted that this construct is informed by the waves of Greek emigration in the late 19th century and in the aftermath of World War II, which was an indirect - as experienced by previous generations or as narrated - experience for the participants of this study.

After the formation of the modern Greek state in the 1830s, Greece witnessed two mass emigration waves. The first wave took place in the late 19th century, when almost a sixth of the Greek resident population emigrated to the USA and Egypt [19]. In 1912 almost 10% of the population had

² Term used by the National Statistics Service of Greece, 2001.

emigrated to the USA, which represents the highest percentage amongst the other European countries since 1900 [20]. Remittances were vital to the local economy, especially as regards rural areas, while the rapid demographic growth encouraged further emigration of the population, which could not be sustained by agricultural economy.

The second wave occurred on the aftermath of the second World War, when more than 1 million of the Greek resident population emigrated to the USA, Germany, Canada, Australia and Belgium [19] [21].

Official statistics show that in the period 1955-1973 Germany absorbed 603,300 Greek migrants, Australia 170,700, the US 124,000, and Canada 80,200 (p. 1, [19]).

It is estimated that in the 1980s, following the restoration of democracy in Greece and the accession into the European Community, almost half the post-war emigrants returned to Greece. Nevertheless the experience of emigration seems to be embedded in Greek history and culture, and it is extensively narrated and researched.

In the process of talking about migration employing the construct of Greek emigrants, participants position themselves ('we'), other Greeks ('the Greeks'), Greek emigrants (either as 'we' or 'they' – the Greeks in Germany, Greek Americans), migrants in Greece ('they' or 'Albanians') and Germany ('they', 'Germans', 'there'). In the extracts that follow reflexive and interactive positioning in the employment of the construct of the Greek emigrant abroad seems to have the following functions; First, the stated or implied analogy between Greek emigrants abroad and migrants in Greece, is used to reinforce an implicit or explicit argument favourable to migrants in Greece. This seems to evolve in a shift in the identity of the latter from a national category to either the (universal) category of the migrant (extracts 1, 3), to foreground the analogy to Greek emigrants, or to the category of the migrant equivalent to the Greek emigrant (extracts 2, 4). The pattern which usually follows is a comparison of Greek emigrants abroad and migrants in Greece on the basis of differences at the expense of migrants in Greece or in favour of Greeks, depending on the emphasis.

In the extracts to follow the analogy between the Greek emigration wave to Germany and the (Albanian) immigration wave to Greece is employed to support the argument that local Greeks should not be prejudiced to the influx of migration since they also witnessed the Greek out-migration wave to Germany, appealing therefore to the right of mobility seeking labour (see Appendix 2 for Transcription Notation).

In extract 1 Costas employs the construct of Greek emigrants abroad to substantiate his argument by building a positive image for migrants in Greece using the analogy to the Greek emigrants abroad.

FG1 - Extract 1

- 96 **COSTAS:** [...] I had a different view about Albanians and about Balkanians (.) I didn't
97 like them very much (.) but through this job I understood that the people just came as
98 migrants >labour migrants< to earn money (.) in order to: raise their children, help their
99 family, help their relatives who are in Albania, in Serbia >I don't know where< >exactly
100 like the Greeks did< who left for America, Australia, Germany
101 **LAMBROS:** [but not
102 **COSTAS:** [they are not criminals

Costas argues that he had a different view about Albanians – a national category – and Balkanians³ – grouping people in the area - in the past in a way also disclaiming that view as an element of the past, which no longer applies (lines 96-97). He retains his distance from these categories in an 'I' and 'they' positioning, while also using a 'mild' structure of negation 'I didn't like them' and leveling 'very much' rather than a strong definitive statement (line 97). The break from this view is introduced

³ Note that this is the term used in the focus group schedule to enable interpretation, thus the participant may have employed it as such.

using 'but' and substantiated by appeal to personal experience, 'through this job' (line 97). In addition, there is a shift from the category of Albanians and Balkanians to the universal category of 'people' who are 'labour migrants' – whom he defines in terms of 'earning money in order to: raise their children, help their family, help their relatives' leading to the example 'exactly like the Greeks did' (lines 97-100). These people 'just came' as labour migrants and not something else 'not criminals' (line 102), foregrounding the argument 'exactly like' Greek emigrants abroad, who were also just labour migrants and not criminals. In this way the mobilization of the analogy of Greek emigrants abroad is used to support an argument of the positive image of migrants in Greece and to explicitly respond on those grounds to the implicit criticism of migrants in Greece being criminals, thus, precluding it before it even appears as an option in the discussion. It should be noted here that reference to crime seems to indicate the shared view amongst Greeks of migrants in Greece engaging in criminal acts.

In the following extract Menelaos mobilizes the construct of Greek emigrants abroad in the context of arguing that he was happy for his Greek national identity when he was younger, which was interrupted by the negative treatment of migrants from 'us' the 'Greeks'. Aggeliki responds to Menelaos' argument by arguing that most migrants coming illegally could justify the attitude of Greeks towards them.

FG 2 - Extract 2

- 776 **MENELAOS:** I let's say was thinking that when the:emigration wave took place tha:t
 777 everyone went to Germany >in the '40s in the '50s around that time< (.) to work (.) and how
 778 many Greeks went to Germany let's say (.) say (..) why then (..) it didn't seem weird to us
 779 suddenly let's say that so many people go to: >Germany to work< and when Albanians came
 780 here suddenly it seems weird to us
 781 [let's say (.) do you understand?
 782 **AGGELIKI:** [yes but
 783 **MENELAOS:** =>you will tell me we didn't create problems etc< I don't know if we did or not
 784 [...]
 785 **AGGELIKI:** I don't agree on this. say in Germany first of all a Greek went after Germany
 786 accepted this Greek (.) if it didn't accept him there was no chance to go (.) while here we say:
 787 **MENELAOS:** yes
 788 **AGGELIKI:** we don't accept them they come illegally (.) we give how much money to turn
 789 them back to Albania, the return illegally (.) and (.) eventually they stay in Greece illegally.
 790 This is it (.) to do what they are doing

The analogy of the Greek emigrant abroad is employed by Menelaos by positioning himself as a local Greek who - along with local Greeks 'us' – shared the experience of observing both the emphasized wave of every Greek going to Germany and of Albanians 'coming' to Greece. The analogy is extended to both events taking place suddenly (lines 779-780). Despite the similarity in terms of observed (and not lived) experience, there is a difference in the attitude of local Greeks as regards Greek emigration – 'not weird' (line 778) and Albanian migration to Greece – 'weird' (line 780). However, no explicit analogy is made to the purpose of migration, which for Greek emigrants abroad was 'to work', whereas no equivalent formulation is offered for the Albanians (lines 777, 779). Thus, while in terms of experience and suddenness an analogy is made, the reasons for the different attitude to the two migration waves is left open. To Aggeliki's 'yes but' Menelaos articulates what he anticipates Aggeliki to imagine (line 783) and positions himself and Aggeliki together with the group of Greek emigrants by using the first person plural (line 783). At the same time he introduces the criterion 'problem', which seems to be a shared one for migrants in Greece, and argues 'we didn't create problems', a negation which implies contrast – 'they' (Albanians) did. He then shifts alignment to speak for himself, as the principal of what is being said but suspends his response to Aggeliki's anticipated criterion on which migration may be considered weird or not by neither agreeing nor disagreeing - 'I don't know if we did or not' (lines 783-784). Aggeliki then uses the singular 'a Greek' to make an objective statement, at the same time not positioning herself as a Greek emigrant,

thus retaining her distance to what is being said (line 785). She also uses a rigid structure – conditional (if clause and result) - and compares the ‘there’, Germany where things were rigid ‘while here’ not that things are not rigid as much but although ‘we’, positioning herself as a local Greek, operate legally on the basis of the accepting/not accepting principle of mobility ‘they’, migrants’ come anyway illegally (lines 788-790). The implicit distinction is that there were no illegal Greek emigrants abroad, not creating problems, no analogy of reception and, therefore, the weird local Greek attitude towards Albanian migrants in Greece is justified. Furthermore, the word ‘illegally’ is used emphatically three times to achieve factuality and also evaluatively precede the profile of migrants in Greece - ‘to do what they are doing’ (line 790). In this way, the construct of Greek emigrants abroad is employed to stress the distinction between the two groups of migrants not so much in favour of Greek emigrants but at the expense of Albanian migrants.

In the following extract Christina mobilizes the pattern – as it is formulated similarly in the focus group discussions - of labour migration to respond to Sofia’s negative presentation of migrants in Greece as not caring about anything but raising money to go and make houses (lines 110-111).

FG 7 - Extract 3

- 106 **CHRISTINA:** do you mean in other words that they don’t care about their relations eh
 107 [with the Greeks?
 108 **SOFIA:** [yes. they care about how to raise money:
 109 **KATERINA:** apart from this
 110 **SOFIA:** thei:r relationship with the Greeks may be seemingly good but their purpose is to
 111 raise money (.) to go and make houses (.) NOTHING else (..) I have talked to A and I have
 112 talked to E (..) and the one works in the store and the other I have her [at home
 113 **CHRISTINA:** [guys every person who
 114 migrates, right, who is a labour migrant comes for a specific purpose he doesn’t go just like
 115 this because
 116 **KATERINA:** [you don’t flatten everything out Christina
 117 **JOANNA:** [we too went to Germany but ((overlapping))
 118 **SOFIA:** the Greeks were not the same when they went to Germany
 119 **CHRISTINA:** that is to say what? Didn’t they have as a purpose to make houses?
 120 **KATERINA:** = you don’t flatten [everything
 121 **SOFIA:** [didn’t they care about the level of thei:r children and the
 122 studies of their children?

Sofia’s argument of migrants in Greece as not caring about anything but raising money to go and make houses (lines 110-111) is strengthened by appeal to personal experience (line 112), which positions her as the author of the point above, not bearing therefore issues of personal interest to a claim which is seemingly strong against migrants in Greece. Christina’s response to Sofia’s argument – notwithstanding its commonsensical theoretical structure on a universally applied definition of migration – is readily met with overlapping speech and the employment of the construct of the Greek emigrant abroad which takes the definition of labour migration on board and introduces a comparison which functions at the expense of migrants in Greece. In this endeavour, Joanna introduces the contrast - ‘we too went to Germany but’ - and positions herself as a member of the group of Greek emigrants abroad by using the first person plural (line 117). This is then continued by Sofia who makes a vague claim about the difference of Greeks, positioning herself outside the group of Greek emigrants abroad, thus achieving distance and objectivity for her argument. Katerina’s reiterated turn, in the second person singular to bring her argument closer to the speaker, stresses that the purpose of labour migration is one thing and reducing everything to nothing – referring to migrants in Greece - which is an extreme, is another. This is continued by Sofia who counters ‘everything’ with the care for children, an ecumenically sensitive topic. Thus, on the one hand migrants in Greece care about ‘NOTHING’ but raising money and making houses while Greek emigrants in Germany cared about the level and studies of their children. This argument is structured in a seemingly rhetorical question using negation to reduce the possibility of a negative response or leveling (lines 121-122). The

function of the contrast between Greek emigrants and migrants in Greece in this extract is to favour the former vis a vis the purpose of labour migration and in so doing reduce the image of the latter in accordance with the negative presentation which preceded this extract.

The following extract is presented to underline the pattern of the dynamic of analogy being readily succeeded by comparison which takes the form of contrast so that the two groups – Greek emigrants and (Albanian) migrants - are not put in the same position. As soon as the analogy is either uttered or implied (see extract 3) – and on the grounds that a positive image is implicitly shared for Greek emigrants – the participants engage in negotiations of difference which, depending on the focus and emphasis, are favourable to Greeks or negative to migrants in Greece.

FG 1 - Extract 4

- 801 **COSTAS:** = we are not examining the extreme phenomena now that they take advantage of
 802 the other. An average employer who takes him to work because he is cheaper labour (.) he
 803 shouldn't say "the Greeks are taking advantage of me"
- 804 **MAY:** ((yes sure))
- 805 **COSTAS:** =the other gives you food, yes
- 806 **FILIO:** in this you are somewhat not right because when the Greeks too went to Germany (.)
 807 even though they went with an invitation (.) >someone had to invite them to go there< (.)
 808 >they went organized with papers doing health tests etc< (.) ε:h the Greeks too were going
 809 there and were saying were swearing on the country where they were going and working (.)
 810 it doesn't mean that because you work because you earn some money you are completely
 811 satisfied [a:nd
 812 **COSTAS:** [I didn't say that you are completely [satisfied
- 813 **LENA:** [yes for what reason for what
 814 reason (.) because if they isolated you from society (.) you go round with the tag "the
 815 Albanian"
- 816 **COSTAS:**= guys there the things are worse let's not compare them (.) [in Germany things are
 817 worse
- 818 **FILIO:** [look in Germany they
 819 went round with the tag of the "Greek" (.) they said "Greek" and it is like we say now
 820 "Albanian" that is to say (.) that is to say an aversion
- 821 **LENA:**= [yes yes
- 822 **COSTAS:** [it was worse there (.) believe me
- 823 **FILIO:** = and even worse (.) because my father was a labour migrant (.) and he put up with
 824 some things >up to a lot I may say< that is to say that they didn't allow them to enter some:
 825 restaurants, that is to say some places where: entrance
 826 was [forbidden
- 827 **LENA:** [yes (.) obviously okay (.) [it is ((like)) dogs entering and milling⁴ here
- 828 **COSTAS:** [there things are much worse (.) we are very lenient here
- 829 **LENA:** =yes okay (...)
- 830 **COSTAS:** =and we are doing well =
- 831 **LENA:** = and in Italy things are much worse (.) in Italy: they take them (.) the Albanians with
 832 the ships (.) they arrive in Brindisi and the others take them and through them a:t Albanian
 833 territorial waters all in (.) they don't accept any Albanian there (.) it's much worse: (.) but
 834 okay (...)

In this extract Costas presents seeking cheap labour as a rational choice, which renders generalised claims by the migrant on the whole of 'Greeks taking advantage of me' groundless (line 803). The shift in footing from principle to author substantiated by active voicing on the part of the migrant in Greece functions to reduce responsibility for what is being uttered. This meets with agreement from

⁴ The connotations of this idiom in Greek would best translate as 'anyone entering and doing whatever they want'.

May which lays the foundation for the escalation of the argument to 'the other', now being the Greek employer who gives 'you', the migrant - closer to the first person 'I' - 'food' which is the basic most important human need and should therefore be highly appreciated (line 805). This could also indicate the attribution of social class which is reserved for migrants in Greece, one which is restricted to satisfying their basic needs. This escalation is picked by Filio who uses the analogy of Greek emigrants abroad to support the argument that one does not have to be completely satisfied just because they work and earn money. Namely, Filio uses the third person plural to position Greek emigrants in Germany, thus not positioning herself as part of this group. She makes an implicit comparison in favour of Greek migrants by using 'even though' and referring to the procedure of Greek migration to Germany, which was strict and systematic entailing invitation, organised papers and health tests (lines 807-808), making also implicit that this was not the case for migrants 'coming' to Greece. After this comparatively positive image, and even if the procedure was organised and systematic, the Greeks still 'swore' at the country (Germany) where they worked. This relatively strong statement is followed by an impersonal structure - 'it doesn't mean' - to reduce agency and achieve objectivity for the argument to follow, which develops as a generic statement on (migrant) satisfaction, retaining the connection to the speaker through the use of the second person singular (lines 810-811). Her statement is strengthened by the extreme levelling of satisfaction - 'completely' (line 811). This lays the ground for Costas to seemingly agree on the level of satisfaction not being complete. Lena picks the analogy, retains the closeness to the speaker by using the second person singular to position the Greek emigrants and 'the Albanian' and offers a reason to support Filio's statement of migrants not necessarily being satisfied just because they work and earn money. The reason is being isolated by society, by Germans - 'they' - and being discriminated against - by local Greeks, thus making the connection with the 'Albanian' in Greece and in so doing acknowledging the shared negative connotation this tag bears (lines 814-815). In this way the analogy of Greek emigrants abroad is used to support the argument of a grounded lack of complete satisfaction on the part of migrants.

The dynamic of this interaction is noteworthy as it develops into a contrast since Costas picks up the reasoning employed by Lena to compare the level of that reason, namely isolation and discrimination. He addresses the discussants with a straightforward call - 'guys' - to draw attention to his argument to follow. He positions himself as speaking on the part of 'here' and counters 'there' to stress the contrast in how different things are between Greece and Germany (lines 816-817). In particular, using the present tense he argues for things being at present worse 'there' beyond comparison. While Costas refers to the present situation, Filio, employs the construct of Greek emigrants abroad to stress the analogy of the past treatment of Greek emigrants by 'they' - Germans - and the present treatment of migrants in Greece by 'we' - local Greeks. By using the first person plural Filio positions herself and the other participants as members of the group of local Greeks, explicitly acknowledging discrimination to Albanians. The analogy of the relationship between Germans and Greek emigrants and local Greeks and Albanians is employed to support the development of the argument presented in lines 801-815, that local conditions interact with migrants' reaction to the host community, and therefore certain behaviours are justified. Interestingly, then, Costas retains the contrast between Germany and Greece - things (conditions) in the former being worse than in the latter - but informed by the previous analogy to the construct of Greek emigrants, he sets his statement in the past (line 822). As his statement 'it was worse there' could be questioned on the grounds of vagueness, Costas makes an explicit plea to be believed by the discussants. This is picked up by Filio who formally breaks the analogy she employed, without, however, losing face since the focus of the narrative she makes by appealing to personal experience is seemingly Greek emigrants in Germany, with no explicit contrast to migrants in Greece. In other words, Filio introduces the comparison by 'even worse', even worse discrimination than the one presented in lines 818-820, even worse than Costas' vague 'worse' conditions. This is rhetorically supported by personal, thus uncontested, experience, as the daughter of a labour migrant in Germany who narrates his experience. It is noteworthy that the segregation presented in lines 823-826, narrates the tolerance of Filio's father as a member of the group of Greek emigrants in Germany, on restrictions from places of socialisation. Lena then shifts positions to compare, using a strong but also anecdotal idiomatic phrase, things 'here', Greece - being very loose, a term commonly used in Greece to refer to lack of organisation or the stereotype of Greek

disorganization (line 827) [22]. As she does not position herself as part of this looseness, she opens the ground for criticism of the regulations pertaining to immigrants in Greece. Costas, retains the present contrast of 'there' – Germany, 'here' – Greece to argue for things being worse in Germany, but shifts the looseness – which has negative connotations – to lenience on the part local Greeks rather than Greece as a state, positioning himself as a local Greek as well (lines 828). While 'very lenient' could also be criticized for the level of lenience being high, thus close to loose, Costas readily supports this as a good practice by 'us', local Greeks (line 830). Thus, being very lenient is a good thing, in contrast to Germany being rigid, which is a bad thing. This line of argument is then further supported by Lena who, retaining the comparison in the present, compares the extreme case of Italy not accepting Albanians (lines 831-834).

To sum up, the employment of the construct of Greek emigrants abroad in this extract functions to initially align Greek emigrants with migrants in Greece in their justified lack of complete satisfaction in the host country. This evolves into a contrast in favour of local Greeks, aligning clearly with lenience 'offered' in Greece than with rigidity 'present' in Germany and Italy. At the same time this functions to undermine the lack of complete satisfaction on the part of migrants in Greece on the basis of a comparative local lenience and exalt the tolerance of Greek emigrants for whom conditions in Germany were much worse.

4. Conclusions

Greece has become a destination country for migrants from Central and Eastern Europe and the former Soviet Union, which taking into account the substantial emigration tradition, informs Greek national identity. In particular, in talking about migration people from Central Northern Greece invoke and mobilize the construct of Greek emigrants abroad to draw similarities with migrants in Greece on the legitimate pattern of labour migration.

Similarities are drawn based on the definition of labour migration in terms of mobility in order to earn and raise money. The analogy evolves in a shift from the restrictive national category/subject of the (Albanian) migrant in Greece to the universal concept of labour migrant or its implicit or explicit equivalent of Greek emigrant. This invocation creates an analogy to the implicit positive image of Greek emigrants, thus, having also positive implications for migrants in Greece. It is usually readily succeeded by comparisons and contrasts in order to identify the differences between Greek emigrants and migrants in Greece. This seems to be employed in order not to retain the two groups in the same (positive) position and evolves in either constructing a comparatively more positive image of Greeks or a more negative image of migrants in Greece, thus returning to the national distinction.

Therefore, it could be concluded that invoking the construct of Greek emigrants abroad is a strategy in the context of talking about migration which functions to create an analogy as regards the legitimacy of labour migration beyond the national subject, to be readily deconstructed on the basis of making national category distinctions at the expense of migrants in Greece or favourable to Greeks.

References

- 1 NSSG (National Statistics Service of Greece) Statistical Yearbook 2001, Athens
- 2 Baldwin-Edwards M. Statistical Data on Immigrants in Greece: an analytic study of available data and recommendations for conformity with European Union Standards, A study conducted for IMEPO by Mediterranean Migration Observatory, UEHR Panteion University 2005
- 3 Willig C. Introducing qualitative research in Psychology: adventures in theory and method. Buckingham: Open University Press 2001
- 4 Stevens R. (eds) Understanding the Self. London: Open University Press & Sage 1996; 222-225
- 5 Gough B. and McFadden M. Critical Social Psychology: An Introduction. New York: Palgrave 2001
- 6 Austin J. L. How to Do Things With Words. Oxford: Clarendon Press 1962

- 7 Carabine J. Unmarried Motherhood 1830-1990: A Genealogical Analysis in Wetherell M., Taylor S., Yates S. J. (eds) *Discourse as data. A guide for analysis*. London: Sage 2001; 268-269
- 8 De Cillia R., Reisigl M., and Wodak R. The discursive construction of national identities, *Discourse and Society* 1999; 10: 149-173
- 9 Billig M. *Banal Nationalism*. London: Sage; 1995
- 10 Billig M. Discursive, Rhetorical and Ideological Messages in Wetherell, M., Taylor, S., Yates, S. J. (eds) *Discourse Theory and Practice: A Reader*, London: Sage 2001
- 11 Galasinska A. and Galasinski D. Discursive strategies for coping with sensitive topics of the other. *Journal of Ethnic and Migration Studies* 2003; 29, 5: 849-863
- 12 Tusting K., Crawshaw, R. and Callen B. 'I know, cos I was there': how residence abroad students use personal experience to legitimate cultural generalizations. *Discourse & Society* 2001;13, 5: 651-672
- 13 Davies B. and Harré R. Positioning: The Discursive Production of Selves in Wetherell M., Taylor S., Yates S. J. (eds), *Discourse Theory and Practice: A Reader*. London: Sage 2001; 140-142
- 14 Burr V. *An Introduction to Social Constructionism*. London: Routledge 1995; 126-127
- 15 Goffman E. *Forms of Talk*. Pennsylvania: University of Pennsylvania Press 1981
- 16 Goffman E. On face-work: an analysis of ritual elements in social interaction. In: *Interaction Ritual: Essays on Face-to-Face Behaviour*. New York: Pantheon Books 1967; 5-45.
- 17 Edley N. Analysing Masculinity: Interpretative Repertoires, Ideological Dilemmas and Subject Positions in Wetherell M., Taylor S., Yates S. J. (eds) *Discourse as data. A guide for analysis*. London: Sage 2001; 198
- 18 Heritage J. Goffman, Garfinkel and Conversation Analysis in Wetherell M., Taylor S., Yates S. J. (eds) *Discourse Theory and Practice: A Reader*. London: Sage 2001; 50
- 19 Kasimis C. and Kasimi C. *Greece: A History of Migration*. Migration Policy Institute. <http://www.migrationinformation.org> June 2004; 1
- 20 Mazower M. *The Balkans*. London: Weidenfeld & Nicolson 2000
- 21 Gropas R. and Triandafyllidou A. *Migration in Greece at a glance*. ELIAMEP October 2005
- 22 Bozatzis N. *Greek National Identity in Talk: The Rhetorical Articulation of an Ideological Dilemma*. Unpublished Ph.D. thesis. Lancaster: Lancaster University 1999

Appendix A

Rhetorical Strategies: Definition

The first rhetorical strategy to be presented is participants' *appeal to Personal Experience*, which constitutes an example on how further coding will proceed in the following rhetorical strategies in the process of mapping strategies to themes and discourses.

Appeal to Personal experience is a common form of argument legitimization. It refers to a narrative of active or passive experience of events which are offered by participants in support or evidence of an unfolding view or argument (see Tusting et al, 2002).

The second rhetorical strategy presented is *Impersonal Structures*. Impersonal Structures consist words, phrases, idioms, sayings, grammar, syntax and hedges which enable the expression of a view or argument in an objective manner. A commonly repeated example is the use of passive voice. This functions to blur agency and disavow accountability by using 'out-there' structures which are not immediately identifiable with the speaker or which exist independently of the speaker. Socially, impersonal structures as explicit mitigators "offer an almost transparent mask of 'political correctness'" (Galasinska and Galasinski, 2003, p. 853).

The third rhetorical strategy presented is *Extreme Case Formulations*. Extreme case formulations consists of referring to examples or making statements which are not mainstream and are stronger than normally expected because they are made in an extreme form. Extreme case formulations are encouraged in focus groups due to the preference for intersubjective agreement which is not as often the case in one-to-one interactions (see Tusting et al, 2002).

The fourth rhetorical strategy presented is *Comparison*. Comparison is a common discursive practice used to understand and classify others based on one's own experience – that being personal and/or social. Beyond the notion of comparison of Social Identity theory⁵, it seems useful to note that comparison becomes analytically relevant in how and when it is being used.

The fifth rhetorical strategy presented is Disclaimer. Disclaimers consist phrases used to disavow agency or mitigate/disclaim or claim objectivity on the position assumed on a point preceding or following.

Finally, *Humor* in the first coding of the transcripts appears to occur to 1. voice strong/extreme views, 2. avoid agreement when a counter position triumphs and 3. to lighten up previously loaded discussion(s).

Appendix B

Transcription Notation

=	no discernible gap between utterances
((text))	researcher's comments
CAPITALS	louder speech
μtextμ	quieter speech
[overlapping speech
<u>Text</u>	emphasized speech
“text”	direct speech
Te::xt	extension of preceding vowel
(.)	short pause
>text<	speeded-up speech
Text*	original term used

⁵ In social identity theory, a social identity is a person's knowledge that he or she belongs to a social category or group (Stets and Burke, 2000). Through a social comparison process, persons who are similar to the self are categorized with the self and are classified as the in-group. Similarly, persons who are different from the self are categorized as different and are classified as the out-group. Categorization, comparison and classification (identification) are recurrent processes of social identity.

EU Police Missions: from Conflict Management to Institution Building in the Western Balkans

Islam Yusufi¹, Andrew Taylor², Miltos Sarigiannidis³

¹SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, iyusufi@seerc.info.

²The University of Sheffield, Dept. of Politics, Elmfield, Northumberland Road, Sheffield S10 2TU, a.j.taylor@sheffield.ac.uk.

³City College, Dept. of Business Administration and Economics, 2 Kalapothaki Str, 54624 Thessaloniki, Greece, msarigiannidis@seerc.info

The paper lays out the scope of the problem of the institutions in the countries of the western Balkans and analyzes the impact of the EU police missions in Bosnia and Macedonia in that regard. Due to the importance of overall work of the European Union in the region, attention is paid to the dynamics that have shaped the work of these missions on the ground. The aim is to assess what has been achieved so far by the missions towards the goal of institution-building; identifying the factors that have shaped the work of these missions in the area of institution-building and factors upon which the success of these missions have depended; and exploring the way forward for these kind of missions. The term *institution-building* has been chosen as it best captures the elements that are there or that should be there in the EU police missions. The criterion for success or failure of the EU missions here in this study is defined as whether these missions have been able to support the host countries in building or nurturing state institutions in instituting or strengthening the rule of law.

Keywords

EU, western Balkans, police missions, institution building.

1. Introduction: Institution-building

Institution-building has been a major feature of the European integration processes since the early days of the European Communities and it became particularly salient with the enlargement of the European Union to the post-communist countries of Central and Eastern Europe. As the European Union embarks upon subsequent enlargements towards the western Balkans where there has been lack of traditions for nationhood and state building, institution-building has emerged as the critical area for the overall work of the European Union's police missions stationed there.

The term *institution-building* refers to building or strengthening state institutions with the aim of establishing or sustaining the system of rule of law that will improve the state's capacity to maintain its security without undermining the human and minority rights of its fellow citizens. In other words, institution-building is strengthening weak or non-functioning states by instituting the necessary rules to tackle the state administration's inability to deliver services, fight organised crime and root out corruption.

Since the inception of the European Security and Defence Policy in 1998 and its operationalisation in 2003, the European Union has invested significant efforts into conducting military, police and rule of law missions in the countries that have recently experienced conflicts or civil unrest. Bosnia and Macedonia are the first countries to host the EU security and defence missions, particularly the police missions EUPM (European Union Police Mission) and Proxima, and the countries that have attracted most of the focus of the EU recently in terms of crisis management and stabilization. In addition, these

police missions are setting standards for crisis management and institution-building in the future endeavours of the EU in the region of the western Balkans or in other parts of the world.

A striking feature of the state governance in the western Balkans is its weakness, particularly in the justice, home affairs and public administration fields. Over-dominance of the state structures in the overall economic and social governance systems is a critical weakness of the countries of the region. In addition, the sources of instability persists, exacerbating the problem of institutions: rising unemployment rates, greater grey economy and growing number of people living under the poverty line; notorious war criminal suspects remaining at large; and organized crime and corruption continuing to be souring business fields.

In addition, the western Balkans states are prone to the organised crime groups and corruption-ridden syndicates that occupy large chunk of government businesses or at least disable the governments' control over them. The most important element of these groups or syndicates is their ability to challenge these governments and disable them in maintaining the rule of law. The regional governments need a well-focused approach with complementary strategy of comprehensive reform that can effectively prevent the weakening of their governance. Recent EU initiatives of establishing security and defence presence in the region, in tandem with the overall European integration agenda, are a positive step in building and strengthening the state institutions in the region.

2. Demand: Institutional deficits

The western Balkan countries of Bosnia and Macedonia experience the calmest period in their post-communist history. International and domestic agreements and arrangements ending the conflicts and confrontations in these countries, including Dayton Peace Accords and Ohrid Framework Agreement and other arrangements, largely are respected and to a great extent implemented, that in turn marks, in principle, the end of an era of post-conflict reconstruction in these countries.

Internally, these countries generally possess democratic governments with a vocation streaming for the Euro-Atlantic integration. The major change that occurred in the region was the downfall of the Milosevic and Tudjman regimes in 2000 that had a strong effect in both countries that strengthened the hand of western oriented and reform-minded political forces as both regimes were pursuing divisive and irredentist objectives, particularly in Bosnia.[1] Only after their replacement, did Bosnia's neighbours begin to work with the rest of the international community to push the country together rather than pull it apart.

Following these changes, what these countries experience is an irreversible process of democratization and moving in the right and positive direction that can lead Bosnia and Macedonia from Dayton and Ohrid implementation, respectively, into the track of accession to EU membership.

Despite these looming positive changes in the region, the weaknesses in public administration, particularly in fighting criminal violence, organized crime and corruption, continuously challenge these countries and hinder their smooth transition from weak states to functioning democratic polities. These weaknesses do not facilitate the establishment of conditions necessary for these two countries to become, as Vachudova puts it:

“unambiguous credible candidates for EU membership” (p: 1, [2]).

Armed insurgents or secessionist movements no longer challenge Bosnia and Macedonia. Both countries earlier passed the “existential test” of their transition; there is no question anymore about the prospect of their future existence. What they face now is the risk of breakdown of law and order in parts of their territories, and the risk coming from the spread of criminal violence, organized crime and corruption that are deeply embedded in their respective national communities.[3] These risks are compounded by the lack of capacity of both countries' home affairs and justice institutions, including police forces, in dealing with those risks. In such cases, what is required is professional, managerial and operational support (read *institution-building*) to the relevant state institutions such as the police forces that in turn will enhance their capacities to deal with complex risks posed by organized crime, criminal violence and corruption. This is not to say that these risks dominate the agenda of these

countries and that they are issues exclusive to them. What is important in these countries is that their law enforcement institutions, prosecutor's offices and judiciary are unable to apprehend, investigate and prosecute the criminals that mainly consist of trans-national smugglers of narcotics, arms, and aliens, traffickers of women for sex, and movers of illicit goods across frontiers.[4]

There is still legacy of the collapse of the Yugoslav state and communist regime in 1989-91 and the subsequent wars that broke the already fragile border defences and controls. Then institutions of accountability disappeared; police standards deteriorated; and centre-local trust was lost, all of which opened up the possibility for the work of these trans-national criminal groups. All contributed to an environment in which organised crime flourished and dislodging or bringing criminal charges against them and prosecuting them currently is difficult due to the absence of necessary normative and operative structures in the countries' rule of law systems that will make possible credible criminal investigation and the prosecution of crimes committed.

More lives were lost in criminal violence in Macedonia in the months after the end of the conflict than during the heat of the conflict itself. In the country there are a number of organised crime groups, estimated up to 10, with an average of 10 members each.[5] Most of them tend to be part of large international networks of smugglers of aliens, narcotics, arms, cigarettes, sugar, cars and others. In October-December 2004, Kondovo, a suburb of Skopje, Macedonia's capital, became a no-go area for Macedonian state security institutions. The suburb came under the control of an armed group expressing dissatisfaction with the political processes in the country. The issue was resolved after a quiet and long effort of the domestic political forces.

If Bosnia's police forces were well trained, equipped with necessary tools, under the clear civilian oversight structures coordinated at the state level reflecting the overall multiethnic character of the state, they currently would be in position to tackle resistance to the return of refugees, and to prevent cross border smuggling, particularly of human beings, drugs and arms that continue to be soaring business fields in the country.

Thus, weak governance compounded with weak police forces is problematic as they lack capacity to deal with the forces that are able to turn a country into breeding ground for organised crime and criminal violence that can affect regional stability and well being of wider region. Thus, the western Balkan countries Bosnia and Macedonia have been in need of support that will strengthen their governments' ability to institute the rule of law, and to maintain secure borders and monopoly over the means of coercion. However, home affairs and justice reforms in the transitional countries such as Bosnia and Macedonia are a long-term processes and they can be destabilizing in short term. Thus, it is the role of the European Union integration process that can play a complementary role as the "hole opened at the pressure pot preventing explosion."

This is the "demand" side, but what has been the "supply" of the European Union taking into account the weakness of the governance in the region? It would be well to consider the policies of the EU towards the western Balkans.

3. Supply: EU policy; Stabilisation or institution-building?

The European Union has both a political and a security interest in preventing failure of governance in the region of western Balkans: a *political interest* in successfully rounding up the map of future of the Union together with the countries of the western Balkans; and *security interest* in stabilising the continent's neighbourhood that has been one of the sources of destabilisation in the continent.

Currently, there are strengthened relationships between the western Balkan countries and the EU. EU's summit in Thessaloniki in June 2003 reaffirmed and confirmed the concrete terms and partnerships for the EU membership perspective of the western Balkan countries that previously were echoed at the Feira and Copenhagen European Councils of the Union. The Thessaloniki Summit reiterated EU's determination to fully and effectively support the European perspective of the western Balkan countries and it adopted European Partnerships, which aim at further strengthening the relations between the EU and the Western Balkans drawing on the previous enlargement experience

when similar Partnerships were offered to the candidates from Central Europe in 1998. With the European Partnerships, the Stabilisation and Association Process was enriched and it was agreed that the Process would remain the framework for the European course of the Western Balkan countries all the way to their future accession.[6] This led to a new era in the region's transition processes that is linked with the preparation of the countries of the region for staging negotiations for EU membership.

Parallel to this development there is a gradual restructuring of the international presence in the region. UN and NATO, international organizations that were involved in the first stages of crisis management in the region in early 1990s, with the turn of century, they gradually are pulling out of the region with the European Union set to increase its role in the region. This is the result of a number of factors. Four points suffice:

- The launch of the Stabilization and Association Process in 1999 offered EU membership perspective to the western Balkan countries;
- Adoption of the European Security and Defence Policy and its operationalisation with the deployment of the first ever EU military and police missions in the region;
- European Union becoming the region's main trading and investment partner; and
- The Union playing a major role in various security arrangements. The Ohrid Framework Agreement of August 2001, and the end of the constitutional deadlock in Serbia and Montenegro being the most evident examples.

All of these demonstrate the growing role of the European Union in the regional affairs. Also, the security challenges of the countries of the region shift from maintaining peace to tackling organized crime and dealing with the governance and state institutions.

In sum, the EU's policy towards the region of the western Balkans consists of two main pillars: EU membership perspective; and the presence of the Union's security and defence missions. The police missions in Bosnia (EU Police Mission) and in Macedonia (Proxima), with the European police officers monitoring, mentoring, and advising their local colleagues to help them improve standards and strengthen the rule of law, form an important part of the EU approach towards the region. Putting in place a practical approach tackling the problem of work of institutions that will make them capable to control the territory, safeguarding the rights of the citizens and laying the foundation for promoting economic and social development has been complementary tool of the EU in moving the countries from stabilisation to association with the EU. Reforming police institutions safeguarding the security of fellow citizens is not sufficient for economic and social development and in turn European integration, but without them neither can be achieved.

4. EU's Security Entrepreneurship: Putting in practice the European Security and Defence Policy

Putting the European Security and Defence Policy into practice in the region of the western Balkans, is one of the areas where the European Union has made the most rapid and spectacular progress. It took just four years to develop the European Security and Defence Policy from vision (St. Malo meeting of France and Great Britain in 1998) to reality (launch of European Union Police Mission in Bosnia in 2003). This has been a result of a successful work of relevant EU institutions to establish a framework for autonomous action of the EU with the aim to contribute to the maintenance and to the promotion of peace and stability in its neighbourhood and wider.

Police is an area, which has received the most attention and is an area in which capacity development has progressed fastest among all priority areas of the EU in strengthening of its civilian crisis management capabilities set at the Feira European Council of June 2000. Member states within the agreed time framework provided 5,000 policemen to cover the range of police operations from advisory, assistance and training to executive missions and committed a rapid reaction capability of 1,400 policemen to be available within 30 days.

On January 1st, 2003 the Union launched a police mission in Bosnia (European Union Police Mission – EUPM) that took over the UN-led International Police Task Force mandated since the Dayton/Paris peace accords of December 1995 to establish law enforcement institutions in Bosnia. EUPM is first ever operation within the framework of European Security and Defence Policy. EUPM is mandated with non-executive function (unlike UN mission which was executive) with the political objective in helping and strengthening (rather than substituting) Bosnian police authorities in establishing sustainable policing arrangements under local ownership in accordance with best European standards and international practices.

Following the launch of EUPM, EU continued with its commitment for putting in practice its common security and defence policy and this time embarked upon its first ever-military mission. Military mission Concordia, launched in March 2003 took over NATO-led Allied Harmony mission, with the aim to contribute to the stabilization and confidence building in Macedonia, a country that narrowly escaped the possibility of a civil war in 2001.

Concordia was a success, but as the mission progressed, greater importance needed to be given to the fight against organised crime and corruption. It was progress in these fields that was essential for assuring the rule of law and re-establishing confidence to state institutions. Macedonia's situation demonstrated that the EU's policies had to change accordingly. In the new realities of Macedonia, it became more relevant to deal with police matters rather than with matters of defence character. As the main threat to stability of the country was no longer armed conflict but criminality, the EU support had to focus on civilian and not on military instruments.

With the launch of Concordia and its work, the Union became aware of the challenges on the ground and redefined its presence accordingly that led to overall shift of its overall policy from stabilisation to institution-building that had effects to all missions of the European Union in the region. The change and shift in EU's policies towards Macedonia was crowned with the launch of a new police mission Proxima on 15 December 2003, which became a tool in fully establishing a rule of law around the country.

Consisted initially of 200 and later of 140 policemen, with additional local staff, Proxima's members are co-located around the country alongside police commanders at medium and senior levels. Some of the mission's members are co-located inside the structures of the Ministry of Interior, including at the Unit for Security and Counterintelligence, supporting the senior police officials in dealing with various policing issues, including tackling some of the serious crimes. Mission members advise police officials in how to conduct raids and operations; they take part in the planning phase of them; observe, and following the end of the operations, they conduct a review of the work done by tackling the deficiencies observed during the planning and conduct of the operations. The mission has also co-located some of its members at the borders with Kosovo, Albania and Serbia, building confidence among the police and prosecutors across borders.

5. General Observations

The EU Police Mission in Bosnia and police mission Proxima in Macedonia, as the first civilian crisis management operations undertaken under the EU's security and defence policy, are an important learning experience for the host countries and for the EU. For the EU, they are a test of one of its main tool, *institution-building*, in the course of conduct of support to the law enforcement institutions damaged in the crises. They provide an important test for the EU capacity to plan and conduct missions according to the local needs. As such, they have helped EU to rectify some very considerable deficiencies in its security and defence policy. With these police missions EU has also made success to overcome the obstacles that have stand on its way to pressure for substantial reforms at EU level that would turn relatively developing area of European Security and Defence Policy into an area where EU can become more proactive.

The way ahead for western Balkan countries is the institution-building, particularly in the field of law enforcement that is critical in circumstances like that of the western Balkans where several nations and nationalities have to coexist within one state framework. Police with structures that transcend nations

and overcomes nationalism is building block of this effort. The only solution to the existing differences in the region is to seek through organising and strengthening the respective states that can administer with all the citizens respecting the differences among the ethnicities. In this context, the role of police, in particular, and law enforcement institutions, as a whole, can play a bridging function as it can facilitate the respect for differences on the ground. Thus, another key lesson of the EU police missions in the western Balkans has been the importance of developing local capabilities at first.

References

- 1 Jacques Rupnik, "The Demise of Balkan nationalisms? A sceptical view," in *The Western Balkans moving on*, (ed) Judy Batt, EUISS, Chaillot Paper No. 70, October 2004, Paris, p. 99.
- 2 Milada Anna Vachudova, "The EU, the Balkans and Turkey: Can 'Soft Power' Bring Stability and Democracy," Woodrow Wilson International Centre for Scholars, Meeting Report 280, Washington DC, 14 October 2003, p. 1.
- 3 Christopher A. Corpora & Louisa I. Shelley, "Tragedy, Transition, and Transformation: The local international nexus of transnational organized crime in the former Yugoslav republics," Woodrow Wilson International Centre for Scholars, Meeting Report No. 252, p. 1.
- 4 David Binder, "Organized Crime in Balkans," Woodrow Wilson International Centre for Scholars, Meeting Report No. 269, 22 January 2002, p. 1.
- 5 Federica Curtol & Ernesto U. Savona, "Country profile: Macedonia," in *The Contribution of Data Exchange Systems to the Fight Against Organised Crime in the SEE Countries*, Final Report, Transcrime, November 2004, p. 358.
- 6 Council of the EU, Thessaloniki European Council, 19 and 20 June 2003, Presidency Conclusions.

SESSION 4:

**PSYCHOLOGY
AND CULTURE**

Visuospatial Attention in Dyslexia

Margarita Diakou¹, Ana B. Vivas², Rod I. Nicolson³

¹SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, mdiakou@seerc.info

²City College, Dept. of Psychology, 24 Pr. Koromila st, Thessaloniki 54622, Greece, vivas@city.academic.gr

³The University of Sheffield, Dept. of Psychology, Western Bank, Sheffield, S10 2TN, UK, r.nicolson@shef.ac.uk

The present paper is an introduction to our study of visuospatial attention in dyslexia, employing the inhibition of return paradigm. The major current theories of dyslexia such as the phonological, the magnocellular, and the cerebellar hypotheses are briefly discussed in terms of their suitability to explain the deficits observed in this syndrome. More specific issues on attentional function in dyslexia are then addressed, with an emphasis on the experimental paradigm developed to assess its function. Finally, we outlined future research that is needed to investigate the validity of the main theoretical accounts of dyslexia.

Keywords

Dyslexia, visuospatial attention, inhibition of return.

1. Introduction

Developmental dyslexia is commonly defined as a specific reading and writing disability occurring in children with normal intelligence and conventional schooling and socio-economic opportunities, in the absence of apparent sensory or emotional problems. Typically, dyslexic children's reading level falls two years behind that expected given their chronological age and years of education. In terms of cognitive function, they appear to have a great difficulty in phonological processing, namely in associating a letter symbol to its corresponding sound [1]. Many studies, however, have shown that in addition to the specific language impairment, these children may have other cognitive deficits related with visual and auditory processing [2-3], attention [4], speed of information processing [5], or motor skills and balance [6]. Thus, developmental dyslexia has been better defined as a syndrome rather than as a single unidimensional disorder.

According to the European Dyslexia Association, dyslexia affects 8% of the population worldwide, independently of language or culture and, as it is not a disease, it cannot be cured. However, with early diagnosis and appropriate intervention, children and their families are better able to cope with the problems originating from the condition. Modern western societies emphasize the need for education, and consequently, children who are not able to learn as efficiently as others, within the structure of the school curriculum, may become stigmatized as lazy or even dumb and experience negative feelings of failure and inadequacy. Thus, special education programs and support for both children and their parents are necessary in order to enable dyslexic individuals to reach their full potential. Although Leonardo da Vinci, Albert Einstein or Niels Bohr became great scientists despite being dyslexics, recent research suggests that the number of dyslexics among prison inmates is at least three times as large as among the general population [7]. The relation between criminality and dyslexia may be explained as a result of social exclusion due to academic failure when there is no support from the educational system.

Early intervention and effective treatment are important for dyslexic children, however, in order to develop useful tools, the ontogenesis, the underlying causes of the disorder need to be identified.

Currently, there are several theories accounting for dyslexia, but none of them seems to fully account for all the deficits observed in people with dyslexia.

2. Theories of developmental dyslexia

2.1 The phonological deficit theory

This is the most widely accepted among the theories of dyslexia and the most thoroughly researched. The phonological deficit theory postulates that the causal deficit in dyslexia is phonological in nature [1]. That is, phonology, the component of language concerned with the rules governing the structure and sequence of speech sounds, is disrupted in at least two dimensions: ‘phonological segmentation’, breaking speech (sentences or words) into its constituent parts (e.g. dog: d-o-g) and ‘phonological awareness’, the ability to associate a symbol (grapheme) with its correspondent sound (phoneme). As a result, dyslexic individuals acquire deficient phonological representations and phoneme-grapheme mapping and, therefore, develop poor phonological and reading skills [8]. In a recent study with dyslexic university students, Ramus and colleagues [8] demonstrated that the phonological deficit was persistent into adulthood and was sufficient to cause dyslexia in the absence of other symptoms in the auditory, visual, or motor domains. Currently there is a growing body of findings from neuroimaging studies with dyslexics in an attempt to identify its neural basis in the context of the phonological deficit hypothesis. Results, though, are difficult to interpret for several reasons. Firstly, the brain networks involved in language and reading are widespread and the role of each structure is difficult to be defined. In addition to this, brain plasticity further complicates the interpretation of the deficits, as hyperactivity of one area may well reflect the brain’s effort to compensate for the original deficit. Moreover, small differences between tasks can result in large differences in performance. Leaving aside these difficulties, research has suggested that dyslexia is linked with increased activation of inferior frontal areas associated with phonological processing [9], reduced activation of parietal and temporal regions also associated with phonological processing and rhyming [9-10] and lack of activation of the insula, which connects the posterior and anterior brain areas [10]. Some authors have proposed that the cerebral dysfunction in dyslexia is mainly localized in the areas associated with phonology and reading surrounding the temporoparietal junction [11]. The parietal cortex and the temporoparietal junction have been related to visual attention and do not seem to be specifically involved in phonological processing and reading [12]. However, the proponents of the strong version of the phonological deficit hypothesis argue that the dysfunction is purely phonological in nature. This contradicts the complexity of the behavioral manifestations observed in dyslexia. Thus other theories have been put forward in order to accommodate the heterogeneity of the dyslexia syndrome.

2.2 The sensory deficit theory

2.2.1 Auditory rapid processing deficit

Tallal and collaborators [3] argued that the primary deficit of developmental dyslexia is not phonological in nature, but instead the phonological problems observed reflect a defective auditory input. In their study, they found that dyslexic children were impaired at discriminating differences in the frequency or the intensity of brief or rapidly presented auditory stimuli [13]. Indeed, dyslexic children have difficulties in discriminating between perceptually similar phonemes, such as ba/pa or da/ta. According to the auditory rapid processing deficit theory, their cognitive system fails to build solid phonological representations, which ultimately impairs reading and writing. Originally this hypothesis did not propose a neuropathological correlate of the cognitive deficit, but this issue was later on addressed by the crossmodal magnocellular account [2]. An obvious weakness of the auditory rapid processing deficit theory is that it, too, fails to provide an explanation for the visual, attentional, and motor problems often observed in dyslexic individuals.

2.2.2 Magnocellular deficit

Given the importance of vision in reading, early researchers in the field of dyslexia had termed this disorder “word blindness”. It has been suggested that the human visual system operates via two major visual channels, the transient, stimulated by low spatial frequencies, and the sustained, activated by high spatial frequencies [14]. More recently, though, as visual processing has been better elucidated, the functions of the transient and the sustained visual channels have been proposed to correspond to the neuronal properties of two distinct visual pathways; the magnocellular (M) and the parvocellular (P), respectively [15]. The M pathway responds faster and is mainly concerned with object localization (“where” system) whereas the P pathway responds more slowly and is specialized in the processing of fine details, essential for object identification (“what” system). Moreover, the two pathways differ in their cortical target areas. Thus the P pathway mainly projects to the ventral stream in the inferior temporal cortex, and the M pathway largely projects to the dorsal stream in the posterior parietal cortex [16].

Stein and Walsh [3] proposed that the magnocellular system is selectively affected in dyslexia. In line with this, it has been found that dyslexics exhibit reduced sensitivity to high temporal frequencies, low luminance, and low spatial frequencies [17]. Also, anatomical studies have reported abnormalities in the magnocellular layers of the lateral geniculate nucleus of the thalamus in this population [18]. Moreover, Stein and Walsh [3] suggested that the magnocellular dysfunction is not specific to the visual domain but generalizes to the auditory and tactile modalities as well, thus providing a neuroanatomical substrate for the auditory rapid processing hypothesis and forming a comprehensive theoretical framework for the diverse manifestations of the dyslexia syndrome. Yet, it has been argued that dyslexics are not impaired exclusively in auditory and visual tasks that tap the magnocellular system [19]. In addition to this, it is not clear how a magnocellular dysfunction may account for reading deficits, as reading should depend on the function of the parvocellular system, responsible for the identification of letters. One of the proposed explanations for this paradox postulates that visuospatial attention might play a mediating role in reading, by enhancing individual letter processing in a sequential fashion (smooth movement of the attentional spotlight) [20]. Steinman and collaborators [21] found that visual attention is predominantly guided by cues activating the magnocellular pathway. It is, therefore, assumed that defective magnocellular input would affect parietal function, thus leading to poor control of spatial attention, resulting in target mislocalisation (e.g., in the case of reading, letter mislocalisation). However, further research is needed to elucidate this issue.

2.3 The cerebellar deficit theory

Automatisation is the process by which, in time, skills become familiar and can be performed without conscious effort [22]. According to Nicolson and Fawcett [23], it is exactly this ability that is compromised in dyslexia. As a result, dyslexics fail to become fluent in relatively simple tasks, but are usually able to compensate by trying harder. Their impairment becomes evident when faced with more demanding tasks, as more resources are consumed in carrying out each of their simpler component parts and central capacity is not sufficient for successful execution of the whole. The present hypothesis posits that the deficit in automatisisation is a general one and that reading difficulties reflect its manifestation in the phonological domain. Taking into account motor and balance impairments as well as difficulties in skill acquisition and time estimation often observed in dyslexics, the authors [24] proposed that a possible neuroanatomical locus of the disorder might be the cerebellum. Data from anatomical and brain imaging studies, which revealed differences in activation and structure of the cerebellum of dyslexic individuals as compared to that of normal readers, further supported this theory [25-26]. Traditionally, the function of the cerebellum has been linked to motor control and only recently its contribution to higher-order cognitive functions has been put forward [27]. Nicolson and collaborators [24] pointed out that there exists a direct link between abnormal cerebellar function and balance impairment and lack of skill automatisisation in dyslexia, as the role of the cerebellum in motor control and skill acquisition has long been established. In addition to this, the authors proposed that poor writing is the outcome of defective motor control. Importantly, in the Nicolson and colleagues’

model [24], problems of motor control also lead to non-optimal articulatory skills, which in turn disrupt the function of the phonological loop. The phonological loop is the working memory subsystem responsible for the operation of phonological analysis [28]. Its dysfunction leads to disrupted phonological awareness, which, as already mentioned, lies at the core of the phonological difficulties constitutional to dyslexia. In general, the cerebellar deficit theory seems to bring together, and explain, the disparate problems that comprise the dyslexia syndrome. In light of the above discussion, it appears that there exists a significant amount of disagreement among researchers in the field of dyslexia. Recently, however, there has been a trend to merge existing theories, in an effort to better understand the disorder and develop more useful tools for its diagnosis and treatment.

All theorists agree that the main problem of dyslexics at the behavioural level is phonological; the underlying cognitive and neural mechanisms, though, are still a matter of controversy. One major reason for this is the nature of the disorder per se, namely its heterogeneity. There have been efforts to classify dyslexics into subtypes depending on the type of symptoms they exhibit and it is possible that different subtypes have different neural underpinnings [24]. A second problem is the significant comorbidity of dyslexia with other developmental disorders, which makes it difficult to differentiate and assign symptoms to disorder [8]. A third issue, inherent to the study of all developmental neuropsychological disorders, is that what is observed at the time of diagnosis is not the direct result of the primary neurological deficit, but rather the outcome of the interaction of primary damage with brain plasticity, the ability of the brain to compensate for the dysfunction [29]. In addition to this, environmental factors, such as supportive or non-supportive parents and existence or lack of proper care, and individual differences, such as overall intelligence or persistence, can help shape the individual profile of dyslexics and highlight or mask the problems. It is clear that the earlier the diagnosis and the intervention the better the outcome. Dyslexia, however, cannot be formally diagnosed before the child has had two years of schooling, so it would be useful to identify markers of the disorder before that age. One of the cognitive functions that can be assessed before children learn to write and read is attention. In the next section issues related to attention and dyslexia will be discussed.

3. Visual attention in dyslexia

It is generally agreed upon that the human cognitive system is of limited capacity. Attention is the cognitive process that selects information for further processing, thus enabling the coordination of behavior in order to achieve a goal. Selection occurs on the basis of the information's relative importance, for instance its novelty or its accordance to current behavioural goals, enhancing cognitive flexibility. Contrary to other domain-specific cognitive processes, attention appears to operate in a domain-general manner [30]. Several cognitive models for attention have been proposed, but in their seminal 1990 paper, Posner and Petersen [31] identified three attentional functions, which are prominent across different models: sustained attention, selective attention for conscious processing, and vigilance. A more recent approach based on converging lines of evidence from behavioural and neuroimaging studies suggest that the attentional system can be divided into at least three separate subsystems performing distinct functions: executive control, visuospatial attention, and alertness. Not surprisingly, given the heterogeneity of dyslexia and its significant overlap with attention deficit disorder (ADD) and attention deficit with hyperactivity disorder (ADHD) [32], dyslexics exhibit impaired performance in various attentional tasks pertaining all three aspects of attention. It has been found that reading-disabled children exhibit more interference than normal readers in a Stroop task, which reflects executive functioning [33]. Also, they are slower to respond to the no-cue condition in a Posner task, which has been interpreted as reduced alertness [34]. The literature on visuospatial attention and dyslexia is quite extensive and has produced controversial results. As visuospatial attention is a complex cognitive function, it is useful to clarify some relevant theoretical issues.

The term visuospatial attention refers to the process of bringing into focal vision objects for more efficient processing. It has been proposed that three cognitive operations underlie the orienting of attention to space: *disengagement* of attention from the currently attended location, *movement* of attention, and *engagement* to the new location, so that the target can be detected and a response can be

made [35]. Each particular operation is thought to be subserved by a different brain area of the posterior attentional network. Positron emission tomography and neuropsychological studies indicated that the posterior parietal lobe is involved in the disengagement of attention [36-37], the superior colliculus is involved in the moving process [35], and thalamic areas are involved in engaging attention [38]. Although orienting is usually achieved by eye- and/or head movements towards targets (overt orienting), it has been found that areas in the visual field can be attended to without -or prior to- eye movements (covert orienting) [39]. To study orienting of attention, Posner [39] developed the cuing paradigm, according to which participants were required to respond to the onset of a target in their visual field, while maintaining central fixation. Prior to the target one of two kinds of spatial cues was presented: a central cue, an arrow indicating the location of the upcoming target, or a peripheral cue, a brief illumination of one of a number of box outlines, directly marking the location of the target. A control condition was included, with a “plus” sign presented centrally, indicating that a target was about to appear, while revealing no information about its location. The results showed that, in a simple detection task, participants responded faster and more accurately to targets appearing at a previously cued location and that this facilitation effect followed both central and peripheral cues. Conversely, when cue validity was manipulated and the cues were misleading, it was found that only the central cues could be ignored, while attention was drawn involuntarily towards the location marked by peripheral cues. Posner [39] suggested that there should be two different mechanisms underlying orienting of attention: an endogenous system, under participants’ intentional control, and an exogenous system, which is controlled by the relative saliency of the sensory input. He hypothesized that these systems were subserved by different neural mechanisms, thus introducing the notion of componential attention. However, recent neuroimaging studies indicated that endogenous and exogenous cues activate similar frontal and parietal areas, suggesting that overlapping neural circuits may mediate voluntary and automatic orienting of visuospatial attention [30].

In a series of studies examining visuospatial attention in dyslexia Riita Hari and collaborators [40-41] found that dyslexics exhibited a right-sided spatial bias in selecting and processing visual information, in addition to a general “sluggish attentional capture”, consistent with mild right parietal dysfunction. Taken together, these findings suggested that dyslexic individuals might exhibit signs of hemi-neglect, and, as suggested by the authors, an attentional deficit could be at the basis of the disorder. Spatial neglect is a disorder usually caused by right parieto-temporal lesions and is characterized by severe inability to orient to the contralateral hemispace [42]. Hari and colleagues [41] suggested that defective magnocellular input to the parietal lobe constitutes the basis for the behavioural manifestations of the disorder. In this line of thought, Facoetti and collaborators [43-44, 6] investigated the allocation of visuospatial attention in dyslexia using variations of the Posner task. While initial results were in support of the mini-neglect hypothesis, later findings were inconsistent. Moreover, they focused mainly in studying the facilitatory effects of spatial cueing on performance, whereas the equally important inhibitory effects remained unexplored. Thus, the systematic investigation of visuospatial inhibitory attentional effects will be the basis of our present research.

Early facilitation of responses to targets appearing at previously cued locations has been well established in literature and has been attributed to automatic (exogenous) orienting of attention towards the cue [45]. However, it has been found that when the peripheral cues are uninformative, and the cue-target interval is longer than 300ms, the initial facilitatory effect is reversed and participants are slower to respond to targets appearing at previously cued locations [46]. This latter inhibitory effect is termed inhibition of return (IOR) [46] and has been found to be a robust phenomenon. Its nature, though, in other words whether IOR should be considered a motor or an attentional phenomenon is still under debate [47, 45]. Early studies of the anatomical structures involved in the development of this effect indicated that it was mediated by midbrain structures associated with saccadic eye movements (e.g., the superior colliculus) [35]. Posner and colleagues [35] found that patients with progressive supranuclear palsy, affecting the superior colliculus, did not exhibit IOR effects in a detection task, while neurological patients with frontal lobe damage, Parkinson’s disease, and parietal lesions exhibited a normal IOR effect. It was concluded that IOR must reflect a bias in oculomotor programming, suggesting that the nature of IOR could be a motor bias rather than an attentional effect [45]. However, recent findings from parietal [48-49] and neglect patients [50]

demonstrated that the function of the posterior parietal cortex might be responsible for the IOR effect, offering support for the attentional nature of the phenomenon.

Inhibition of return has been consistently replicated in detection tasks, however some early studies investigating whether the effect would be present in more complex tasks failed to obtain confirmatory evidence, thus questioning the robustness of the phenomenon across tasks. Further research, though, indicated that inhibition of return is observed in discrimination tasks, albeit following a different time course [51]. It has been suggested that IOR takes longer to develop and is faster to dissipate as a function of task difficulty [45]. Other studies have shown that the time IOR first appears may depend on the duration of the peripheral cue [52] and brightness and spatial position of the cue [53]. In a recent study examining the time course of inhibition of return as a function of task difficulty, the target discriminability was manipulated within tasks by using degraded stimuli as targets. The results indicated that IOR appeared later in both detection and discrimination tasks as the targets became more difficult to discriminate [54]. A similar design will be employed in our study in order to examine IOR in relation to dyslexia.

4. The present study

4.1 Design

The first part of the present study will focus on examining IOR in dyslexic individuals using the paradigm developed by Posner [39]. A series of four experiments will be administered:

4.1.1 Experiment 1a

In Experiment 1a, we are going to investigate the allocation and temporal resolution of covert visuospatial attention in a detection task using the Posner paradigm. A single cue procedure is going to be administered within blocks having four different cue-target intervals (150, 300, 500, 900ms SOA). The experiment will start with a cross will appearing at the screen center for 500ms, followed by the initial display of the three boxes (one at the centre, one left and one right) presented for 1000ms. Then one of the peripheral boxes will flicker for 45ms followed by an interstimulus interval (ISI; the three boxes display) for 105, 255, 455 and 855ms, according to the four different SOAs. Then the target will appear inside either the cued (valid) or the uncued (invalid) peripheral box and remain onscreen until a response is made or for 2000ms. The next trial will begin with the central cross at fixation display. The target will be either an intact or a degraded stimulus with equal probability. Experimental trials will be intermingled with catch trials, where no target will be presented, to control for anticipatory responses.

The within-subjects factors will be hemifield (left or right), location (cued or uncued), type of target (intact or degraded) and SOA (150, 300, 500, and 900ms).

4.1.2 Experiment 1b

In Experiment 1b, a double-cue procedure will be employed. After the 45ms peripheral cue, a 30ms ISI will follow and then the central box will flicker for 45ms followed by a second ISI of 30, 180, 380, and 780ms, according to the four different SOAs. Then the target will appear. The remaining procedure and the factors will be the same as in Experiment 1a.

4.1.3 Experiment 2a

In Experiment 2a a single-cue discrimination task will be employed, and participants will be asked to press one of two keys depending on the identity of the target (square or circle). The remaining procedure and factors will be the same as in Experiment 1a.

4.1.4 Experiment 2b

In Experiment 2b a double-cue discrimination task will be used, while the remaining procedure and factors will be identical to Experiment 1b.

4.2 Participants

The study will employ three groups, one experimental and two control groups, one matched for chronological age and one for reading age. The age range for the experimental and the chronological age control groups will be from 9 years (fourth grade of elementary school with three years of schooling) and 12 years (sixth grade with five years of schooling). The reading age control group will be selected on the basis of the dyslexics' performance on the reading test. Participants will be recruited from Thessaloniki public or private elementary schools.

Before deciding whether a child meets the criteria for entering the study, a battery of tests will be administered. First, children will be assessed with a short version of the WISC-III, which includes Similarities, Arithmetic, Picture Completion and Block Design [55]. The short version is preferred because it can accurately predict scores obtained by the full-scale version, without being time consuming and tiring, as it only lasts 27 minutes. Then, the Test for Reading Performance (TORP) [56] will be administered. The children will undergo brief fixation training, developed with the E-Prime [57] in order to get accustomed with the procedure. The training will vary in duration depending on successful performance. Finally, participants will be administered the Attentional Network Test for Children (ANT-C) [58]. The ANT-C is an experimental paradigm testing the efficiency of attentional networks and is not yet standardized; therefore scores of dyslexics will be compared with scores obtained by control groups. Finally, the experimental procedure will begin.

4.3 Predictions

As described earlier, there are several theories accounting for dyslexia. Each proposes different causal routes and different neuropathologies associated with the disorder. It is interesting to see what is predicted for IOR in detection and discrimination tasks using intact and degraded stimuli from these different perspectives.

4.3.1 The phonological deficit theory

According to the phonological deficit theory no attentional problems are anticipated, since dyslexia is conceptualized solely as a phonological disorder. Therefore dyslexic children are not expected to differ from chronological or reading age controls in measures of visuospatial attention, and if any differences are observed they should be attributed to comorbidity. Even the view that postulates the possible locus of pathology is the temporoparietal junction, does not imply a causal relationship between attentional problems originating from the dysfunctional area, but rather argues they must be comorbid.

4.3.2 The magnocellular deficit theory

Within the visual modality, the M cells are responsible for the processing of transient and peripheral visual events and their output mainly feeds into the dorsal (where) visual stream, which carries out the function of target localization, also by directing the allocation of visuospatial attention [21]. In the present experiment, peripheral cues are transient visual events, which are supposed to activate the M cells, thus enhancing attentional selection. It is generally agreed upon that visual selective attention is at least partly determined by stimulus salience as well as current behavioural goals [59]. If the M cell response is not optimal and the input to the dorsal visual stream is defective, the relative salience of cues for dyslexic individuals could be reduced and not result in effective attentional capture. This hypothesis could be directly tested in a subsequent experiment by manipulating cue duration and therefore its relative salience in order to establish a threshold for cue duration to achieve optimal attentional capture. In the present experiment, however, if attention is summoned in the periphery by the cue, and it is anticipated that it will, as in previous experiments [4] peripheral cues with duration as short as 40ms elicited cue effects, dyslexics are expected not to differ from controls in terms of performance of target detection and discrimination. According to the mini-neglect hypothesis, general slowness in responding, reduced alertness and problems in sustained attention are expected, but the strongest prediction of this hypothesis is that attentional deficits should be lateralized. Hence it is expected that, under both task conditions, dyslexics should be slower to respond to targets appearing

in the left visual field, especially following invalid cues, and that IOR for targets appearing in the right visual field should be reduced in agreement with data from studies with neglect and parietal patients [48-50]. Previous studies with dyslexic individuals were inconsistent in finding lateralised deficits and there are no data for IOR, since a double cue procedure has never been used.

4.3.3 The cerebellar deficit theory

According to the automaticity deficit theory, performance in detecting targets presented in the periphery following an uninformative cue should be equal for both hemifields and it is expected that dyslexics should be overall slower than controls. However, as the detection task should be more difficult in terms of attentional demands for dyslexic individuals, IOR should appear in both hemifields, but later than in controls, as the time course of IOR has been found to be a function of task difficulty. Moreover, if dyslexics employ more attentional resources to the detection task, the magnitude of the effect should be greater than that of controls, as it has been suggested that the more attention is originally allocated the greater the IOR [45]. In addition to this, in the discrimination task, IOR should follow the same time course in both the intact and degraded target conditions [60] in the dyslexic group, while controls should exhibit IOR later in the degraded condition as compared to the intact [54].

Regarding the contribution of the cerebellum to attentional shifting, there are two contradictory lines of thought; i) Ivry and colleagues [61] suggest that the cerebellum is not directly involved in attentional shifting. ii) On the other hand, Townsend and collaborators [62] postulate an involvement of the cerebellum in visuospatial attention, based on findings from autistic and cerebellar patients. According to their line of thought, delayed responses to targets requiring attentional shifts, such as those following invalid cues, should be expected from dyslexic individuals, as they are known to exhibit cerebellar abnormalities. It is expected that in the single cue condition dyslexics should be faster to respond to targets presented at the cued location at longer SOAs in contrast to normal controls, who should exhibit facilitation at short cue-target intervals. Moreover, if the mechanism responsible for the generation of IOR operates after attention is allocated and removed from the periphery, IOR is expected to occur later in dyslexics, since facilitation is produced later [45]. If IOR is applied together with facilitation and only manifests itself after benefits fade out, then IOR is expected to appear blunted in dyslexics, as it should be counteracted by delayed facilitation.

5. Conclusions

The study of visuospatial attention in dyslexia appears to be a promising area. Not only it could lead to the development of a tool for the early identification of the disorder, but it can also help elucidate the nature of the reading difficulty. It is clear that more or less all ontogenetic hypotheses anticipate a general slowing in processing speed, an operation closely associated with attention. However, the systematic study of the function of the inhibitory instead of the facilitatory attentional mechanisms may offer support to specific theories, thus adding to the better understanding of the neural underpinnings of the dyslexia syndrome.

References

- 1 Snowling M. Dyslexia: A cognitive developmental perspective. Oxford, UK: Blackwell 1987.
- 2 Stein J, Walsh V. To see but not to read: the magnocellular theory of dyslexia. Trends in Neurosciences 1997; 20; 147-152.
- 3 Tallal P, Miller S, Fitch R H. Neurological basis of speech: A case of the preeminence of temporal processing. Annals of the New York Academy of Sciences; 682; 27-47.
- 4 Facoetti A, Lorusso M L, Cattaneo C, Galli R, Molteni M. Visual and auditory attentional capture are both sluggish in children with developmental dyslexia. Acta Neurobiologiae Experimentalis 2005; 65; 61-72.
- 5 Nicolson R I, Fawcett A J. Dyslexia is more than a phonological disability. Dyslexia 1995; 1: 19-36.

- 6 Fawcett A J, Nicolson, R I. Persistent deficits in motor skill of children with dyslexia. *Journal of Motor Behaviour* 1994; 235-240.
- 7 Kirk J, Reid G. An examination of the relationship between dyslexia and offending in young people and the implications for the training system. *Dyslexia* 2001; 7(2); 77-84.
- 8 Ramus F. Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology* 2003; 13; 212-218.
- 9 Shaywitz S E, Shaywitz B A, Pugh K R, Fullbright R K, Constable R T, Mencl W E, Shankweiler D P, Liberman A M, Skudlarski P, Fletcher J M, Katz L, Marchione K E, Lacadie C, Gatenby C, Gore J C. Functional disruption in the organization of the brain for reading in dyslexia. *Proceedings of the National Academy for Sciences of the USA* 1998; 95; 2636-2641.
- 10 Paulesu E, Frith U, Snowling M, Gallagher A, Morton J, Frackowiak R S J, Frith C. Is developmental dyslexia a disconnection syndrome? Evidence from PET scanning. *Brain* 1996; 119; 143-157.
- 11 Eden G F, Zeffiro T A. Neural systems affected in developmental dyslexia revealed by functional neuroimaging. *Neuron* 1998; 21,279-282.
- 12 Corbetta M, Kinkade J M, Ollinger J M, McAvoy M P, Shulman, G L. Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience* 2000; 3, 292-297.
- 13 MacArthur G M, Bishop D V M. Auditory perceptual processing in people with reading and oral impairments: Current issues and recommendations. *Dyslexia* 2001; 7; 150-170.
- 14 Livingstone M S, Hubel D. Psychophysical evidence for separate channels for the perception of form, color, and movement and depth. *The Journal of Neuroscience* 1987; 7, 3416-3468.
- 15 Milner A D, Goodale M A. *The visual brain in action*. New York: Oxford University Press 1995.
- 16 Ungerleider L G, Mishkin M. Two cortical visual systems. In: Ingle D J, Goodale M A, Mansfield J W, editors. *Analysis of Visual Behaviour*. Cambridge, MA: MIT Press 1982; p. 549-586.
- 17 Borsting E, Ridder W H, Dudeck K, Kelley C, Matsui L, Motoyama J. The presence of a magnocellular deficit depends on the type of dyslexia. *Vision Research* 1996; 36; 1047-1053.
- 18 Livingstone M S, Rosen G D, Drislane F W, Galaburda A M. Physiological and anatomical evidence for a magnocellular defect in developmental dyslexia. *Proceedings of the National Academy of Sciences of the USA* 1991; 88; 7943-7947.
- 19 Amitay S, Ben-Yehudah G, Banai K, Ahissar M. Disabled readers suffer from visual and auditory impairments but not from a specific magnocellular deficit. *Brain* 2002; 124; 2272-2285.
- 20 Vidyasagar T R. Neural underpinnings of dyslexia as a disorder of visuospatial attention. *Clinical and Experimental Optometry* 2004; 1; 4-10.
- 21 Steinman B A, Steinman S B, Lehmkuhle S. Transient visual attention is dominated by the magnocellular system. *Vision Research* 1997; 37; 17-23.
- 22 Schneider W, Schiffman R M. Controlled and automatic human information processing: 1. Detection, search, and attention. *Psychological Review* 1977; 84; 1-66.
- 23 Nicolson R J, Fawcett A J. Automaticity: a new framework for dyslexia research? *Cognition* 1990; 35;159-182.
- 24 Nicolson R J, Fawcett A J, Dean P. Developmental dyslexia: The cerebellar deficit hypothesis. *Trends in Neurosciences* 2001; 24; 508-511.
- 25 Rae C, Lee M A, Dixon R M, Blamire A M, Thompson C H, Styles P, Talcott J, Richardson J A, Stein J F. Metabolic abnormalities in developmental dyslexia detected by 1H magnetic resonance spectroscopy. *Lancet* 1998; 351; 1849-1852.
- 26 Eckert M A, Leonard C M, Richards T L, Aylward E H, Thomson J, Berninger V W. Anatomical correlates of dyslexia: Frontal and cerebellar findings. *Brain* 2003; 126; 482-494.
- 27 Akshoomof N A, Courchesne E, Press G A, Iragui V. Contribution of the cerebellum to neuropsychological functioning: Evidence from a case of cerebellar degenerative disorder. *Neuropsychologia* 1992; 30, 315-328.
- 28 Baddeley A D. *Essentials of Human Memory*. East Sussex, UK: Psychology Press 1999.
- 29 Thomas A, Karmiloff-Smith M. Are developmental disorders like cases of adult brain damage? Implications from connectionist modeling. *The Behavioral and Brain Sciences* 2002; 6; 727-750.
- 30 Fernandez-Duque D, Posner M I. Brain imaging of attentional networks in normal and pathological states. *Journal of Clinical and Experimental Psychology* 2001; 23; 74-93.

- 31 Posner M I, Petersen S E. The attention system of the human brain. *Annual Review of Neuroscience* 1990; 13; 25-42.
- 32 Ramus F. Neurobiology of dyslexia: A reinterpretation of the data. *Trends in Neurosciences* 2004; 27; 720-726.
- 33 Everatt J, Warner J, Miles T R, Thomson M E. The incidence of Stroop interference in dyslexia. *Dyslexia* 1999; 3; 222-228.
- 34 Heiervang E, Hugdahl K. Impaired visual attention in children with dyslexia. *Journal of Learning Disabilities* 2003; 36; 68-73.
- 35 Posner M I, Rafal R D, Choate L S, Vaughn J. Inhibition of return: Neural basis and function. *Cognitive Neuropsychology* 1985; 2; 211-238.
- 36 Parasuraman R, Greenwood P M. Selective attention in aging and dementia. In: Parasuraman R, editor. *The Attentive Brain*; Cambridge, MA: MIT Press 1998; p.711-733.
- 37 Rafal R. Neglect. In: Parasuraman R, editor. *The Attentive Brain*; Cambridge, MA: MIT Press 1998; p. 489-546.
- 38 Arrington C M, Carr T H, Mayer A R, Rao S M. Neural mechanisms of visual attention: Object-based selection of a region in space. *Journal of Cognitive Neuroscience* 2000; 12; Supplement 2; 106-117.
- 39 Posner M I. Orienting of attention. *Quarterly Journal of experimental Psychology* 1980; 32; 3-25.
- 40 Hari R, Valta M, Uutela K. Prolonged attentional dwell time in dyslexic adults. *Neuroscience letters* 1999; 271; 202-204.
- 41 Hari R, Renvall H, Tanskanen T. Left mini-neglect in dyslexic adults. *Brain* 2001; 12; 1373-1380.
- 42 Kerkhoff G. Spatial hemineglect in humans. *Progress in Neurobiology* 2001; 63; 1-27.
- 43 Facoetti A, Paganoni P, Turatto M, Marzola V, Mascetti G G. Visual-spatial attention in developmental dyslexia. *Cortex* 2000; 36; 109-123.
- 44 Facoetti A, Turatto M, Lorusso M L, Mascetti G G. Orienting of visual attention in dyslexia: evidence for asymmetric control of attention. *Experimental Brain Research* 2001; 138; 46-53.
- 45 Klein R M. Inhibition of return. *Trends in Cognitive Sciences* 2000; 4; 138-147.
- 46 Posner M I, Cohen Y A. Components of visual orienting. In: Bouma H, Bouhuis D G, editors. *Attention and Performance X*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc 1984; p. 513-556.
- 47 Kingstone A, Pratt J. Inhibition of return is composed of attentional and oculomotor processes. *Perception and Psychophysics* 1999; 61; 1046-1054.
- 48 Vivas A B, Humphreys G W, Fuentes L J. Inhibitory processing following damage to the parietal lobe. *Neuropsychologia* 2003; 41; 1531-1540.
- 49 Vivas A B, Humphreys G W, Fuentes L J. Abnormal inhibition of return: A review and new data on patients with parietal lobe damage. *Cognitive Neuropsychology: Special issue on inhibition of return* 2006
- 50 Bartolomeo P, Scieroff E, Decaix C, Chokron S. Modulating the attentional bias in unilateral neglect: the effects of strategic set. *Experimental Brain Research* 2001; 137; 432-444.
- 51 Lupianez J, Milan E G, Tornay F J, Madrid E, Tudela P. Does IOR occur in discrimination tasks? Yes it does, but later. *Perception and Psychophysics* 1997; 59; 1241-1254.
- 52 Collie A, Maruff P, Yucef M, Danckert J, Currie J. Spatiotemporal distribution of facilitation and inhibition of return arising from the reflexive orienting of covert attention. *Journal of Experimental Psychology: Human Perception and Performance* 2000; 26; 1733-1745.
- 53 Pratt J, Hillis J, Gold J. The effects of physical characteristics of cues and targets on facilitation and inhibition. *Psychonomic Bulletin and Review* 2001; 8; 489-495.
- 54 Castel A D, Pratt J, Chasteen A L, Scialfa C T. Examining task difficulty on the time course of inhibition of return: Detecting perceptually degraded targets. *Canadian Journal of Experimental Psychology* 2005; 59; 90-98.
- 55 Kaufman A S, Kaufman J C, Balgopal R, McLean J E. Comparison of three WISC-III short forms: Weighing psychometric, clinical, and practical factors. *Journal of Clinical Child Psychology* 1996; 25; 97-105.
- 56 Panteliadu S, Sideridis G D. Discriminant validation of the Test of Reading Performance (TORP) for identifying children at risk of reading difficulties. *European Journal of Psychological Assessment* 2000; 16; 139-146.

- 58 Fan J, McCandliss B D, Sommer T, Raz A, Posner M I. Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience* 2002; 14; 340-347.
- 57 Schneider W, Eschman A, Zuccoloto A. E-Prime. Pittsburgh: Psychology Software Tools Inc 2001.
- 59 van Zoest W, Donk M. the effects of salience in saccadic target selection. *Visual Cognition* 2005; 12; 353-375.
- 60 Moores E, Nicolson R I, Fawcett A J. Attentional deficits in dyslexia: Evidence for an automatisisation deficit? *European Journal of Cognitive Psychology* 2003; 15; 321-348.
- 61 Ravizza S M, Ivry R B. Comparison of the basal ganglia and cerebellum in shifting attention. *Journal of Cognitive Neuroscience* 2001; 13; 285-297.
- 62 Townsend J, Courchesne E, Covington J, Westerfield M, Singer Harris N, Lyden P, Lowry T P, Press G A. Spatial attention deficits in patients with acquired and developmental cerebellar abnormality. *The Journal of Neuroscience* 1999; 19; 5632-5643.

Autism, Communication Impairment, and Gastrointestinal Symptoms

Sophia Dimopoulou¹, Efrosini Kalyva², Michael Siegal³

¹SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, sdimopoulou@seerc.info

²City College, Dept. of Psychology, 24 Prox. Koromila Str, 54622 Thessaloniki, Greece, kalyva@city.academic.gr

³The University of Sheffield, Dept. of Psychology, Western Bank, Sheffield, S10 2TN, UK, m.siegal@sheffield.ac.uk

Autism is a developmental disorder characterized by impaired communicative development [1]. Some children with autism have a delayed or total lack of development of speech. In other cases, they fail to respond to others' speech (for example, the child does not orient to his own name). Children with autism often use stereotyped and repetitive use of language (for example, they show pronoun reversal; saying 'You' for 'I'). They make idiosyncratic use of words and appearance of neologisms and at the same time they fail to initiate or even sustain conversation normally. Abnormalities of prosody (pitch, stress) also underline the deficits in communication in autism. Such difficulties in communication present challenges to parents of children with autism. In particular, they often need to cope with their children's gastrointestinal symptoms such as chronic diarrhoea, abdominal pain, night-time weaking and irritability [2]. According to previous studies [3] [4], even many typically developing children as young as 3 and 4 years of age have rudimentary understanding of contamination in that they recognize that a substance that looks edible can be harmful in reality. However, little is known about the ability of children with autism to understand the nature of contamination even though a substantial number of children with autism face gastrointestinal problems that may possibly be more severe than those of typically developing children. The aim of this review is to underline the need to investigate, interrelationships between the ability of children with autism to judge the goodness of food, contamination sensitivity and gastrointestinal symptoms.

Keywords

Autism, communication, gastrointestinal symptoms.

1. Communication in Autism

Considerable interest has been shown in the understanding of the nature of language and communication impairments that are among the primary characteristics of autism [5]. Children with autism exhibit delayed or impaired language development, they also experience persistent abnormalities of pragmatic language or the use of language in meaningful contexts. These pragmatic difficulties are evident in conversation [6], understanding jokes [7], figurative expression [8], and inferencing [9].

These pragmatic difficulties are attributed to deficits in theory of mind (ToM) understanding [10]. According to the theory of mind hypothesis, individuals with autism have profound difficulty in interpreting a person's action within a mentalistic framework. This type of impairment involves deficits in social-affective relatedness to other people, beginning during the first year of life, followed by a more advanced cognitive deficit in representing attitudes or the content of an individual's mental state, which emerges during the preschool years. Thus, from very early children with autism fail to understand other's behavior that could be due to the failure to orient social stimuli [5]. This deficit has

an impact on impairment in communication, because the essential motivation to communicate lies in the desire to share intentions, thoughts and emotions with other individuals.

The theory of mind hypothesis has generated a great amount of new research on language and communication that has been underlined by the significant role of mental state understanding in acquiring language. Different types of deficits are found among children with autism [11]. A first deficit is found in the asymmetry in preverbal communication. While children with autism use gestures or vocalization to express their needs they will not communicate objects of shared interest. Such protodeclaratives that involve joint attention entail an incipient understanding of intentionality [12]. When children with autism show signs as to acquire spoken language they do not express a full range of language functions or speech acts. Children with autism do not use language to share information with others, or to ask for new information; rather the limited use of language to exchange information, which reflects the preverbal communicative deficits in joint attention and protodeclarative gestures, reflects the lack of children's with autism in understanding that people might know different things and that language is the key means for discovering the contents of another individual's mind [13].

Because children with autism show restrict use of language primarily for instrumental functions, their conversational abilities also remain rather limited. They do not make progress in initiating, expanding or even elaborating on a topic, despite the relatively good vocabulary and syntax [14]. Nevertheless, conversational turn-taking abilities are not totally impaired and the desire to communicate might partly explain echolalia, stereotyped language and repetitiveness that is often found in children with autism spectrum disorder [10].

Knowledge that the speaker and the hearer have different perspectives and roles is another aspect of language that is impaired in autism spectrum disorder. Once more, this reveals the lack of mentalistic understanding about people's unique perspectives that is underlined in the errors that children with autism make in reversing personal pronouns such as referring to themselves as 'you'. Even among those individuals with autism who acquire functional language, reflecting their theory of mind, specific impairments remain [14].

These deficits extend across both the verbal and the non- verbal aspects of communication. At the verbal level, problems with narrative discourse, involving impoverished stories, difficulties in understanding and using nonliteral language and providing the appropriate level of relevant information could be quite striking. At the non-verbal level, the prosody and the paucity of gesture that are being described even in individuals with high functioning autism, suggest difficulty conveying emotional attitudes [13].

The deficits in communication are encompassed by core impairments in understanding mental states and are universal across the range of autism spectrum disorders. However, it is important to note that the range of expression varies from one individual to another. Among individuals we can view developmental changes in communicative deficits. Almost all children with autism show reverse pronouns when acquiring language and rely on echolalia and stereotyped language.

Variation in communicative deficits is related to the severity of the disorder, specifically to the theory of mind and executive function impairments. Studies on language in autism underline the continuous nature of impairments in the theory of mind and suggest that the close relationship that exists between theory of mind and language might provide promising methods of remediation [10].

2. Do Children with Autism show more Gastrointestinal Problems than do Typically Developing Children?

Such difficulties in communication present challenges to parents of children with autism. In particular, they often need to cope with their children's gastrointestinal symptoms such as abdominal pain that may be greater than those in typically developing children and they may not be as sensitive to contaminated substances that can result in such symptoms

Horvath and his colleagues [15] investigated the presence of gastrointestinal abnormalities in children with autism spectrum disorders. In their study, children with autism underwent medical examinations in order to examine whether there is a relationship between autism disorder and gastrointestinal symptoms. The findings suggested a significant relationship between autism and the appearance of gastrointestinal symptoms.

Molloy and Manning- Courtney [2] also found children with autism spectrum disorder facing chronic gastrointestinal symptoms (diarrhea; bloating; abdominal pain; night-weakening or unexplained irritability) and that each child had at least one chronic gastrointestinal symptom. However, such findings are based on medical records and have no association with using reasoning tasks (false belief tasks of executive function tasks).

Black and her colleagues [16] indicated that there was no evidence that children with autism were more likely than typically developing children to have had gastrointestinal symptoms at any time before their diagnosis of autism. According to the methodology of this study, case reports were used in order to identify and described the gastrointestinal symptoms among children with autism, without using any type of false belief tasks in order to identify the ability of children with autism to judge the goodness of food.

By contrast, Jyonouchi and his colleagues [17] also highlighted the relationship between gastrointestinal symptoms and dietary intervention in young children with autism. They used medical records and innate immune responses as measured by the production of proinflammatory and counter-regulatory cytokines. The findings of the study suggested a possible link between gastrointestinal symptoms mediated by innate immune abnormalities.

It could be concluded that previous studies have identified that the majority of children with autism face gastrointestinal symptoms, using only medical examinations and case reports from hospitals. Nevertheless this type of investigation should go deeper by being examined in a more cognitive level and not strictly to a medical one.

3. Contamination Sensitivity in Autism

According to previous studies [4] [3], even many typically developing children as young as 3 and 4 years of age have rudimentary understanding of contamination in that they recognize that a substance that looks edible can be harmful in reality. However, little is known about the ability of children with autism to understand the nature of contamination even though substantial number of children with autism face gastrointestinal problems that may be more severe than those of typically developing children.

Contamination sensitivity is usually underlined by a failure to perceive that a substance could be transformed to a harmless one by contact with a foreign object. Thus, young children could judge the goodness of food only by its phenomenal appearance [18] [19]. In a sense, this understanding requires knowledge similar to that in ToM tasks in which reality does not correspond to beliefs based on appearance that may be false.

Contamination is illustrated by rejection of good food that has been in contact with a disgusting entity (contaminant). This effect is viewed as an illustration of the sympathetic magical law of contagion [20] [21]. This law underlines that when two things make contact, their properties are being exchanged and they could be affected. It has been demonstrated that contagion was operative in the beliefs and attitudes of typically developing adults [22] and in their rejection of food that has been in contact with a disgusting entity. In this study, typically developing adults showed some instances of contagion beliefs after being provided with a certain questionnaire.

Although children are used to reject prototypical disgust entities by the time they are two to three years old, the contagion feature of disgust will be absent until the age of seven. Fallon and her colleagues [23] have presented to children, aged between four and ten year of age, illustrated stories. In each of the stories, they had a different negative object (contaminant) that was falling into a favorite beverage. Then the beverage was de-contaminated; the offending entity was removed; then the

beverage was spilled and the glass was refilled with fresh beverage and then the glass was washed three times before being refilled. The findings of the study indicated that children showed their willingness to drink the beverage at each stage by pointing to one of a series of cartoon faces, varying from happy to sad in appearance. Although children aged between six to seven years dislike the disgust entity and usually did not like the beverage when the contaminant was in the beverage, they did like the beverage when the contaminant was removed. These types of results were also confirmed by other studies [24] [25] [26].

Contagion should not be regarded as a primitive or even an infantile idea. On the contrary, it is suggested that contagion is a sophisticated belief. Usually there is no visible or sensory record of contagion; rather it is the history of an entity that determines certain attitudes. Contagion is based on the perception that appearance is not equal to reality [3]. As an articulated or a false belief, contagion might require conception of invisible entities that are exchanged during the contact time. Thus, if contagion is viewed as an articulated or a false belief then the need of having a Theory of Mind considers to be necessary.

However, some methodological issues have risen from the limited number of techniques that have been used in order to investigate children's conceptions of contamination sensitivity [20]. The effects of social pressure in such studies are difficult to evaluate. Sometimes, children might have misunderstood the requirements of the particular situation; in other words they might know that a drink that has been in contact with a disgust entity could be harmful, nevertheless they might not be aware that an adult could offer them a contaminated drink in an effort to test their understanding of the consequences of contamination sensitivity (gastrointestinal symptoms).

Future research should investigate the ability of children with autism to judge the goodness of food in examining the relation between contamination sensitivity, ToM reasoning, and gastrointestinal symptoms. In that case it could be argued that children with autism will be more likely than children without autism to display contamination sensitivity, and that this sensitivity can be seen in terms of the extent to which they display ToM reasoning and the extent to which they have been reported to have symptoms of gastrointestinal problems.

References

- 1 Frith, U. Autism: Explaining the enigma. Oxford: Blackwell: 2003.
- 2 Molloy C A, & Manning-Courtney P. Prevalence of chronic gastrointestinal symptoms in children with autism and autistic spectrum disorder. *Autism* 2003; 7: 165-171.
- 3 Siegal M, & Share D L. Contamination sensitivity in young children. *Developmental Psychology* 1990; 26: 455-458.
- 4 Hejmadi A, Rozin, P, & Siegal M. Once in contact, always in contact: Contagious essence and conceptions of purification in American and Hindu Indian children. *Developmental Psychology* 2004; 40: 467-476.
- 5 Norbury C F. The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology* 2005; 23: 383-399.
- 6 Hale C M, & Tager-Flusberg H. Brief report: The relationship between discourse deficits and autism symptomatology. *Journal of Autism and Developmental Disorders* 2005; 4: 519-524.
- 7 Sigman M, & McGovern C W. Improvement in cognitive and language skills from preschool to adolescence in autism. *Journal of Autism and Developmental Disorders* 2005; 35: 15-23.
- 8 Charman T. Matching preschool children with autism spectrum disorders and comparison children for language ability: Methodological challenges. *Journal of Autism and Developmental Disorders* 2004; 34: 59-64.
- 9 Dennis M, Lazenby A L, & Lockyer L. Inferential language in high function children with autism. *Journal of Autism and Developmental Disorders* 2001; 31: 47, 54.
- 10 Wilkinson K M. Profiles of language and communication skills in autism. *Mental Retardation and Developmental Disabilities Research Review* 1998; 4: 73-79.

- 11 Howlin P. Outcome in high functioning adults with autism with and without early language delays: Implications for the differentiation between autism and Asperger syndrome. *Journal of Autism and Developmental Disorders* 2003; 33: 3-13.
- 12 Bruinsma Y, Koegel R L, & Koegel L K. Joint attention and children with autism: A review of the literature. *Mental Retardation and Developmental Disabilities Research_Review* 2004; 10: 169-175.
- 13 Happe F G E. Understanding minds and metaphors: Insights from the study of figurative language in autism. *Metaphor and Symbolic Activity* 1995; 10: 275-295.
- 14 Mundy P, & Markus J. On the nature of communication impairment in autism. *Mental Retardation and Developmental Disabilities Research Review* 1997; 3: 343-349.
- 15 Horvath K, Papadimitriou C, Rabsztyrn C, Drachenberg S, & Tildon T. Gastrointestinal abnormalities in children with autistic disorder. *Journal of Pediatrics* 1999; 5: 1-8.
- 16 Black C, Kaye J A, & Jick H. Relation of childhood gastrointestinal disorders to autism: Nested case-control study using data from the UK general practice research database. *BMJ* 2002; 325: 419-421.
- 17 Jyonouchi H, Geng L, Ruby A, & Zimmerman-Bier B. Dysregulated innate immune responses in young children with autism spectrum disorders: Their relationship to gastrointestinal symptoms and dietary intervention. *Neuropsychobiology* 2005; 51: 77-85.
- 18 Rosen A B, & Rozin P. Now you see it, now you do not: The preschool child's conception of invisible particles in the context of dissolving. *Developmental Psychology* 1993; 29: 300-311.
- 19 Springer K, & Amy B. The role of physical contact and association in early contamination sensitivity. *Developmental Psychology* 1994; 30: 864-868.
- 20 Rozin P. Social and moral aspects of eating. In Rock (ED.), *The legacy of Solomon Asch: Essays in cognition and social psychology*. Potomac, MC: Erlbaum 1990; 97-110.
- 21 Rozin P, Haidit J, & McCauley C. Individual differences in disgust sensitivity: Comparisons and evaluations of paper-and-pencil versus behavioral measures. *Journal of_Research in Personality* 1999; 33: 330-351.
- 22 Rozin P, Nemeroff C, Wane M, & Sherrod A. Operation of the sympathetic magical law of contagion in interpersonal attitudes among Americans. *Bulletin of the Psychonomic_Society* 1989; 27: 367-370.
- 23 Fallon A E, Rozin, P, & Pliner P. Development of food rejection with special reference to disgust and contamination sensitivity. *Child Development* 1984; 55: 566-575.
- 24 Rozin P, Millman L, & Nemeroff C. Operation of the laws of sympathetic magic in disgust and other domains. *Journal of Personality and Social Psychology* 1986; 50: 703-712.
- 25 Siegal M. Children's knowledge of contagion and contamination as causes of illness. *Child Development* 1988; 59: 1353-1359.
- 26 Rozin P, Fallon A E, & Augustoni-Ziskind M. The child's conception of food: The development of contamination sensitivity to "disgusting" substances. *Developmental Psychology* 1985; 21: 1075-1079.

Consuming Nationalism: Contemporary Nationalism in Turkish Popular Culture

M. Sinan Gökçen¹, Sean Homer², Caroline J. Oates³

¹SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, sgokcen@seerc.info

²City College, Dept. of Business Administration and Economics, 13 Tsimiski Str, 54624 Thessaloniki, Greece, s.homer@city.academic.gr

³The University of Sheffield, Management School, 9 Mappin St, Sheffield, S1 4DT, UK, C.J.Oates@sheffield.ac.uk

Popular culture in Turkey is increasingly communicating nationalist messages since a couple of years. Simultaneously, public opinion polls show declining support for the European Union membership in Turkey and also point out that the population is becoming more receptive to nationalist political propagation. Furthermore, the polls suggest that the voters possess an increasing inclination to support of far-right parties that propagate conservatism and nationalism as their foremost principle. There were a number of reports about nationalist outbursts, such as lynching attempts, mass demonstrations and campaigns ‘reclaiming the national honor’ against so-called ‘internal and external enemies’ of Turkey. Hence, nationalism is evidently the true rising star of the Turkish politics, but an important question to ask is what kind of nationalism is becoming ‘popular’. This study holds that looking into ‘most popular’ popular culture products and profiling the recurrent themes they utilize gives important clues about the contemporary public psyche in Turkey and the particular ‘type’ of nationalism that is evidently finding escalating support among the Turkish people. For this purpose, the messages of nationalism conveyed in three popular culture works that became much well-liked in Turkey, the *Cola-Turka* advertisement campaign, the novel called *Metal Storm* and the TV series-movie named the *Valley of the Wolves* are analyzed in this paper.

Keywords:

Anti-Americanism, EU-Turkey, nationalism, popular culture, Turkey.

1. Introduction

Popular culture in Turkey is using nationalism as a prime source of inspiration for attracting public attention in an ever-increasing manner in the last years. It is obviously not a coincidence that there were frequent outbursts of nationalist manifestations, in which ‘national honor’ was cherished, especially in the 2004-2005 period. These public outcries made nationalism an increasingly discussed issue again in Turkey, principally because of waves of nationalist outbursts that had an unmistakably rancorous tinge. There were intermittent attempts of lynching, politically motivated murders and mass rallies in, all embellished with a nationalist rhetoric [1]. Furthermore, tensions in the south-eastern and eastern regions reappeared, causing frustration among both the Kurds and Turks that the ‘peace times’ were wasted and the spiral of terror started whirling once more [2].

An important question to ask is what kind of nationalism is gaining a foothold in Turkey and why this is occurring. In this regard, the trends of popular culture might give important clues about the contemporary nature of Turkish nationalism and its sources. This paper looks at the ‘most popular’ works of popular culture that specifically targeted conveying nationalist messages for rousing popular interest. It is noteworthy that the most renowned of the popular culture products specifically adopt nationalist rhetoric that emphasizes ‘Turkish superiority’ vis-à-vis the Western culture and power. It is

not only any nationalist message, but the assertion of pre-eminence of the ‘Turks’ conveyed by certain popular culture artifacts that grabs the attention and gains the devotion of the audiences in Turkey. It would not be an exaggeration to state certain advertisements, books, movies and soap operas do capture both hearts and minds of a great number Turkish audiences or readers through reclaiming ‘nationalist pride’ by looking at the great public stir they caused. This paper scrutinizes the cases of a soft drink’s advertisement campaign, *Cola Turka* [3], a novel narrating the eruption of a war between the US and Turkey, *Metal Storm* (Metal Fırtına) [4], and a TV series that was also filmed into a movie, *Valley of the Wolves* (Kurtlar Vadisi) [5]. These three examples are chosen particularly because of the unprecedented public interest they gained among the local viewers. In return, the massive domestic enthusiasm generated international curiosity.

The media, both domestically and internationally, played an important role in turning these three popular culture artifacts into mythical icons of current Turkish nationalism. The recurrent appearance of articles, commentaries, news stories and TV broadcasts regarding these productions led to a snowballing effect in terms of mounting their popularity. As much as the media interest and public attention reinforced each other, the domestic and international political developments also bore a fertile ground for proliferation of nationalism as a preferred theme for popular cultural works in the first place.

2. Rising Nationalism in Turkey

It was unexpected that tide would turn against Turkey when the new millennium started. Materialization of the European Union full-membership perspective into a tangible target at the Helsinki Summit of 1999 cast a transformative spell over Turkey. It was as if foremost problems of the country such as massive corruption amongst politicians and bankers, a network of mafia bosses, repetitive economic crises, continuous human rights violations, and protracted clashes of will with the Kurdish minority evaporated into thin air, and Turkey was converted into a rising star [6]. Optimism was reinforced with the consecutive legislations of constitutional amendments and legal reform packages in 2000-2004, bringing Turkey’s legal framework on a par with the Copenhagen Criteria [7].

However, the path to EU accession has turned into a rather long marathon requiring a level of political and social stamina that Turkey sorely lacks and signs of fatigue have already set in less than half a decade after the Helsinki Summit. By 2005, accusations of ‘treason’, ‘compliance with external enemies of Turkey’ became commonplace, stigmatizing anyone who is working for ameliorating the country’s human rights record, pursuing a revisionist stance towards the traditional narratives of national history, supporting minority rights, and in general striving to turn Turkey into a more open society [8].

The dashing of these initial hopes concerning the future of Turkey derived from both internal and external sources. Domestically, there were number question marks arising with regards to the AKP’s (Adalet ve Kalkınma Partisi - Justice and Development Party) competency as an unwavering and accomplished ruling power dedicated to continuing along with the reform process on the way to the EU membership. Additionally, there have been an increasingly intense struggle for power between secular and Islamist groups as a consequence of the ‘religiously oriented’ policies pursued by the AKP, such as introducing prison penalties for adultery through changing of the civil code and promoting ghettoization of serving alcohol by forming ‘red zones’. For some voters, the AKP came to power specifically as an icon of corruption-free politics and their avowal was marred by several accusations of misconduct. Ideologues of AKP were opting for marrying the concepts of Islamic chastity, conservatism, democratic ideals, humanism and economic efficiency [9]. They appeared to encapsulate the golden political formula of blending an ideal dose of modernity and tradition, but soon news alleging that some of its members are also striving to weave their own web of nepotism and corruption broke [10]. Moreover, the initial pace of reform slowed causing criticism from the EU [11]. At the same time, the EU membership process hastened the proliferation debates regarding ‘taboo’ subjects to the population at large. Among these explosive subjects were the Armenian genocide, the Kurdish Question, the securing of minority rights and the Cyprus issue.

On the international relations front, the War in Iraq and frequent reiteration of the concept of 'clash civilizations', pitting the Muslim world against the West contributed to the sentiments of anomie among the Turkish public. Despite the fact that Turkey has historically been a staunch ally of the United States, anti-Americanism proved to be on the rise due to rampant bitterness about the current US administration's Middle East policies. Furthermore, the threatening attitude of US officials and the heavy pressure applied to Turkey, regarding the authorization for the use of Turkish air space and the moving of American ground troops through Turkish territory into Northern Iraq led to the Grand National Assembly rejecting a bill for complying with US army [12]. Simultaneously, the opinion polls were showing that an overwhelming majority of the Turkish public was against Turkey's involvement in the War in Iraq [13]. However, the mainstream Turkish media was propagating an intensely pro-war stance, and painted a very pessimistic picture of Turkey's future after the bill was rejected (alleging that the US would seek vengeance against Turkey and financial aid would be curbed).

All in all, the stifling atmosphere regarding Turkey's role in the war sparked wide spread resentment among the public. In a couple of weeks, the parliament had to give in to the intense pressure of the US and accepted subsequent bills that opened the way for cooperation with American troops [14]. Public animosity towards the US was provoked to new heights when Turkish soldiers on duty (said to be on a covert mission) in Northern Iraq were taken into custody by the American troops. Claims that sacks were placed on the Turkish soldiers' heads during detention were the key reason for public anger. This confrontation, henceforward called the 'sack' incident in this paper, proved to be lasting in effect. It is still possible to come across references to it, as two of the popular culture works studied here, *Metal Storm* and *Valley of the Wolves Iraq* begun their narrations with this episode.

Meanwhile, the less than supportive attitude of various European countries like Austria and France towards the membership of Turkey has also caused consternation among the Turkish public. This is why public surveys follow a fluctuating pattern regarding support for EU membership. A study conducted in 2002 affirmed that 64% of the people supported EU membership [15]. Another study conducted in 2003-4 found this figure at 75% [16]. Nonetheless, EU support seems to be in decline as a research publicized in May 2006 states that 63,1% of Turkey desires becoming a part of the Union [17].

Nationalism appears as one of the key reasons behind Euro-skepticism. All the three studies concurred that around half of the Turkish public define themselves as 'nationalist'. The most recent pole, Özkırımlı and Uyan's research, place the figure of nationalists among the Turkish people to 52,4% [17]. According to the same study, 50,3% of the Turkish people believe that the EU aims to 'divide' Turkey. The concept of 'dividing' is crucial for understanding the Turkish psyche because of the conflict with the Kurdish population and de facto autonomy of the Northern Iraqi Kurds. Especially after the War in Iraq, there was a wide spread fear that a free Kurdish state will be established. According to Özkırımlı and Uyan's study, the foremost 'fear' of the Turkish public is 'the American threat' by 35% and secondly, an independent Kurdish state by 25,8% . In relation to this trepidation, the prevalent official view that minority rights debates might pose a threat to the integrity of the state finds resonance among some of the public. Through the national education system and propaganda by the state since the foundation of the Republic, the Sèvres Treaty, which was signed by the collapsing Ottoman state and gave the Allied powers tutelage over almost all of Anatolia, is still emblazoned as an imminent threat in Turkish people's minds [18]. According studies, 33,5% to 36% of the Turkish people think that the reforms demanded by the European Union are akin to the Sèvres Treaty's articles [16-17].

Columnists from the mainstream media and liberal academia proposed various analyses for making sense of the change in the air, and many painted a pessimistic picture of Turkey's short-term future. Most commentators predicted that nationalism and extreme right wing political views would gain even further ground [19]. There was also the a tone of dreading about the 'feared' discovery of the 'hidden reality' that Turkey had transformed nationalism into a core social value and it no longer required state propagation to reproduce itself.

The public surveys point out to an increasingly right and far-right oriented voting profile, in which politicization of nationalism is appreciated. At the same time, certain elements of popular culture are embraced as never before, creating fads that bewitch the whole public sphere in Turkey. This is why the analysis of certain specific trends of popular culture between 2003 and 2006 reveals the way in which nationalism has increasingly become a reactive and emotional collective sentiment for the Turkish society. In other words, these trends shed light on the contemporary qualities of Turkish nationalism.

The three cases studied here attracted local media's attention and fascination to the extent that a mutual relationship of benefiting. In that sense, it would not be wrong to assume that initial positive public reaction snowballed into an even greater public interest through the media's extensive coverage. It is not 'news' that the news media has been regenerating nationalist sentiments in Turkey. As media and popular culture are so much intertwined in Turkey as far as propagating capitalizing on nationalism is concerned, here a parentheses must be opened for discussing media's relation with nationalism in general in Turkey.

Newspapers, radios and TV stations in Turkey display 'banal nationalism' in Billig's terms [20]. Applying his theory of everyday nationalism into the case of Turkey, Yumul and Özkırımlı conducted a study of 38 Turkish daily newspapers on the randomly selected day of 16 January 1997 [21]. This day was referred to as an ordinary day as "it was not an official day of celebration, commemoration, electoral campaigning time or a time of extraordinary national crisis" [21]. The results of their quantitative survey depicted how journalists and columnists use nationalist rhetoric spontaneously and report news through nationalist lenses.

Although it can be said that nationalism is a long established reflex of the Turkish media for gaining readership, there is also a simultaneous urge of instrumentalizing nationalist rhetoric as a tool in the assumed 'mission civilatrice' among some mainstream media élite. This urge has a history extending back to the founding days of the Republic and it has to do with the ideal of Westernization projected onto a very traditional mass population.

Naturally, both the content and context of the aforementioned project changed over time. In contemporary times, Bali mentions how the mainstream media élite aims to project the ideal of the 'White Turk' (Beyaz Türk) [22]. This ideal type, the so-called 'new Turk,' is European looking, Mediterranean in culture, physically attractive and well-groomed, adopts as much a refined life-style as possible, is a pioneer in his chosen business and an innovative entrepreneur. He (or she) is modern and embraces technological advances and takes advantage of the globalizing world while remaining loyal to his nation. Bali names Turkey's most circulated daily's editor-in-chief Ertugrul Ozkok as the self-professed 'sculptor' of the White Turk ideal [23]. While the White Turks are ardent supporters the European Union and well-deserved members of the European family, they are also nationalists who "discover the golden formula of the magic of being nationalist without being racist" [22]. Communicating such ideals is not done just for the sake of 'advancing' society, however, as media corporations in Turkey retain strong links with the political parties in power for various business and commercial purposes.

It is against this backdrop that Turkish nationalism is finding expression in the media and the production of popular culture. Especially, the popular culture products display unprecedented degree of nationalist rhetoric and in a way, became mouthpieces for expressing popular sentiments prevalent among the public in general. The cultural artifacts taken under consideration here, the advertisement campaign of the soft drink, *Cola Turka*, the bestselling novel *Metal Storm*, and the soap opera that is also filmed into a blockbuster, the *Valley of the Wolves* became almost iconic because they encapsulated the popular nationalist sentiments. All three became 'internationally' infamous through global media coverage, as well as attracting the interest of scholarly publications. Jingoism is readily traceable in various other elements of popular culture, such as the pop songs, TV magazine and news programs in Turkey, but the chosen artifacts are unique in their extremity in provoking local and international reaction and thus, illustrate Turkish nationalism's contemporary qualities.

3. An advertisement with a Theory: Cola Turka and Positive Nationalism

Since the 1990s, there were numerous advertisements seeking to engender consumers' sympathies through utilizing nationalism in Turkey, but *Cola Turka* outstripped all its precedents by becoming an icon of national defiance [24]. This soft drink is produced by Ülker, one of the largest food and beverage manufacturers of Turkey and rather ironically this company, whose owner is known for his devotion to Islam, was previously boycotted by staunchly secular Kemalists. Therefore, when *Cola Turka* entered the market in 2003 it rapidly became an improbable magnet for attracting nationalist sympathies. This particular product was publicized with perfect timing, as its advertising campaign was launched at precisely the same moment as the controversy surrounding the detention of Turkish soldiers by US troops in Northern Iraq arose. Therefore, the success of the campaign must be seen in the context the turbulent international political atmosphere regarding Iraq and the Middle East as well as the whole saga of deteriorating relations with the US.

The advertisements, starring the US comedian Chevy Chase, communicated the simple message of 'reverse' cultural imperialism by showing how Americans sampling *Cola Turka* became 'Turkish'. The commercial films were shot in New York and basically featured Chase as an 'ordinary' American witnessing his fellow country people and family adopting 'typical' Turkish cultural, national, social traditions upon drinking *Cola Turka*. These stereotypically 'Turkish' behaviors portrayed in the ads were; celebrating a soccer victory draped in Turkish flags, speaking slang in Turkish, eating Turkish food, singing a national hymn, kissing the hands of the elderly and hurling water on the street to wish departing guests a safe trip. In the final shot of the commercial series, Chase sports a black, bushy mustache, symbolizing 'Turkishness' after guzzling *Cola Turka* himself.

The advertisements immediately struck a chord among Turkish audiences, obviously grabbing the viewers' attention and subsequent devotion to the product. Not only did the product itself instantly become the domestic cola market's second best-selling product after *Coca-Cola*, but also it sparked a whole fad of nationalist homage to the splendor of Turkish traditions and the Turkish way of life. The surge of nationalism created by the ads received international publicity from unlikely sources, such as Taiwan, Zambia and Canada [25].

Cola Turka became the symbol of Turkish national pride, which was broken with the soldiers' detention incident, as the timing just coincided. The cover of a popular news magazine branded *Cola Turka* as *Cola-yı Milliye* (Nation's Coke) in allusion to the early 20th century mass movement *Kuva-yı Milliye* (National Oath), the organizing movement behind the war of Independence paving the way for the foundation of the modern Turkish state [26]. There were illustrations encircling the internet which paraphrased the popular nationalist slogan "Either Love here or Leave here", as "Drink it or Leave here" [27]

The campaign's creator, former CEO and Creative Director of Young & Rubicam/ Reklamevi Istanbul, Serdar Erener admitted in an interview that he perceived nationalism as an ideal medium for marketing, as it is an ideology whose star is on the rise. "Yet", he said, his "image of nationalism espouses proud, innovative, optimist and extroverted qualities, in tune with the globalizing world" [28]. Therefore, Erener asserted, he was targeting to capture the imagination of the audiences embracing what he conceptualized as "positive nationalism".

The success of the campaign was obviously coinciding with the infamous 'sack' incident, yet Erener became so confident that he knew the Turkish people well that he wrote an anthropological study on Turkish society, which remains unpublished. Previously, the relation between the advertisement business and promoting nationalism went to other extremes. One of Erener's rivals in the advertisement business, Ali Taran, also prided himself in knowing the Turkish public by heart, but his difference from Erener is taking Turkish people "as they are" without analyzing them or commenting on them [29]. Taran was also very successful in marketing nationalism, to the extent that his promotion campaign re-established an internationally notorious and corrupt businessman Cem Uzan as the promising leader of a new political party within a matter of months and almost led to his election as a parliamentarian [30]. Uzan's party, GP (Genç Parti- Young Party) pursued a campaigning style laden with nationalist motives, and promised the "recuperation of the national honor" [30]. The

instantaneous sensation created by the promotion campaign of GP shows that the Turkish public was ready for nationalist slogans even before the 'sack' incident between the Turkish and the American soldiers, or even the invasion of Iraq, probably due to economic crises and the role of the International Monetary Fund (IMF) in restructuring Turkey's economy, because Uzan frequently mentioned that Turkey was being targeted by the IMF as a country to become 'colonized'. However, it was later on with the invasion of Iraq that nationalism became more pronounced in Turkey.

It is noteworthy that the invasion of Iraq is pivotal for triggering the creation of rivals to the cola-flavored American soft drinks, not just in Turkey, but elsewhere, too. There were other soft drinks than Cola Turka that sought to assert its 'difference' through utilizing a political preference. *Qibla Cola* and *Mecca Cola* were examples of soft drink products seeking to allure Muslim consumers through pursuing them to "liberate their taste" in the case of the former and "to drink with commitment" in the latter [30]. While others advocated charity or capitalization of 'Muslim' magnets instead of foreign ones, *Cola Turka's* case remains unique in the sense that, by the way it asserted cultural superiority. Moreover, cultural superiority cashed, as three years after it is launched, *Cola Turka* now holds over 20% of the market eating up mostly from *Coca Cola's* share [24]. It has also secured a contract with the Turkish Airlines and one of the biggest soccer teams of Turkey, Beşiktaş [31]

In its later advertising campaigns, *Cola Turka* has promoted Turkish military "heroism" as well as promoting football fans. Neither of these campaigns has reached the success of the initial one. The emphasis of the initial campaign was on traditionalism, stressing Turkish culture's superiority and reclaiming a sense of national pride. The later campaigns, however, merely emphasized concepts like "peace in Iraq" and a "gentlemanly approach to football".

Thus, it can be said that what contributed to the earlier campaigns was not "positive nationalism", but nationalism per se. Ironically, *Cola Turka's* first advertisements contained various phrases that combined both English and Turkish, creating a hybrid language while at the same time asserting Turkish culture's superiority. For example, their first commercial film opened with the sentence "New York'ta bir Morning" ("A Morning in New York") flashing over the screen. Another example, of this hybrid language was, when one of the Turkified American character exclaims, "Çok şükür my God!" ("Many thanks my God"). One would have thought that the adulteration of the Turkish language would have insulted domestic audiences, especially when a 'nationalist' message was being conveyed. On the other hand, maybe there was reason for Turks to be proud that 'Westerners' were speaking Turkish in the first place, albeit in a besmirched form.

4. Sweet Taste of Revenge: Metal Storm and Blurring Lines between Reality and Fiction

It can be said that it is rare that literary works astound the whole country in Turkey. In that sense, the domestic and international commotion sparked by the action/thriller *Metal Storm* was really remarkable. A New York Times feature questioned "How many new thrillers have been the subject of testimony before the Senate Foreign Relations Committee? ... (A) Turkey expert from the Nixon Center told the senators that *Metal Storm* was 'essential in understanding the Turkish mind-set today'" [32]. What this book can tell us about the current Turkish state of mind is that it is highly prone to conspiracy theories, anti-Americanism, as well as yearning for national heroes saving the national honor nowadays. However, there is more to the story than this.

The plot of *Metal Storm* is akin to any international thriller of the same genre. There are conspiracies, heroes, foes, treasons, wars, and a lot of villains. The twist lies in the fact that the source of evil is unusual; it is the Americans that jeopardize the world, and it is Turkish heroes that save it. The story commences in 2007, with the real event of the detention of the Turkish soldiers in Northern Iraq portrayed in minutest detail, albeit changing the chronology. Then the fantasy begins; the US infantry attacks Turkish troops on the Northern Iraqi border and this attack is portrayed as if it was initiated by the Turkish army. The US military uses it as a pretext to bomb Ankara and Istanbul by launching

“Operation Metal Storm”. The sudden American viciousness towards Turkey is not coincidental, as the US seeks control of Turkey’s rich borate, uranium and thorium reserves.

The novel’s success derives from its intertwining of fact and fiction, just as many thrillers do. For example, Turkey does possess the majority of the world’s borate reserves and the reserves also possess the potential to become a strategic resource [33]. Moreover, most of the characters of the novel are familiar such as Donald Rumsfeld, Condoleezza Rice, George W. Bush, Tayyip Erdoğan, Abdullah Gül all of whom carrying out their real life duties. After being intrigued into the ‘almost real’ plot, facing ‘real characters’, and becoming emotionally provoked by the detailed atrocities carried out by invading US troops, the reader is ready to believe any further fantasizing. The national culture is being destroyed, the cities are razed to the ground and those escaping to the countryside are waging a “heroic” independence war. This is, of course, highly reminiscent of the classical narrations of the Independence War that led to foundation of Turkey as a nation state after the First World War. Additionally, the name of the second operation launched after “Metal Storm” to “enslave” Turkey is “Sèvres”, and it envisages dividing Turkey between Armenia, Greece and Kurdistan. The US is after much more than the minerals as it turns out. According to the US characters in the novel Anatolia is a historically Christian territory in dire need of being rescued from the ‘barbaric Turks’.

As mentioned previously, the very pronouncement of Sèvres still evokes deep nationalist sentiments in Turkey. Adding to this the new fears of the ‘Crusades’, the ‘Clash of Civilizations’ and the prevalent belief that Turkey is marginalized by the Christians because it has an overwhelmingly Muslim population, the whole storyline pushes the reader to the far ends of nationalist hysteria. In one of the few critiques of the novel to be found in Turkey, Cem Erciyes argues that the book is an utter failure in literary terms, with undeveloped characters that have no credibility or depth. He states that because the characters are so weakly developed, the reader cannot sympathize with them and relate to the story. On the other hand, what the book excels in is inciting national ‘obsessions’ while embellishing Turkey’s history and heroic deeds in between the lines, according to Erciyes [34]. Strikingly, he also asserts that the authors ‘montaged’ nationalist rhetoric into the novel for profiting from the public fears in a “wholly unethical manner”.

Debating Metal Storm’s dubious literary credentials aside, the novel is certainly playing on rallying nationalist feelings in a rather deliberate manner as key themes of Turkish nationalist psyche are repeatedly exploited. On page 27, for example, the Rumsfeld character rejects a plan to open up a military front in the Dardanelles, citing the Western powers defeat by “Turkish troops” during the First World War. The ‘Victory of Dardanelles’ is one of the most celebrated episodes in the Turkish history, with annual fêtes of commemoration staged and its narrations are reproduced both the state and the public through publications and media coverage. Moreover, the nationalities of the villains of the novel are ‘strategically’ chosen as well. For example, among the perpetrators of the invasion of Turkey are, of course, Armenians. Meanwhile, the Kurds ransack state buildings once the US attacks and a Kurdish pilot who is given the privilege of bombing the “father of the Turks”, Atatürk’s grave, cannot stop himself from making a suicidal dive into this most symbolic of nationalist symbols. However, the novel chants, the body of Atatürk has already been stolen, but “his soul is commanding the Independence War”.

Many readers visiting the website of the book’s publisher congratulate the authors on how real the plot was, and yet others indicate that they find it very plausible that the events might very well materialize just as it is foreseen by the authors [35]. Likewise, the authors, Burak Turna and Orkun Uçar, believe that they fashioned not just a work of fiction, but from their point of view, “it is a philosophical and scientific calculation”, and they are convinced that a war between the US and Turkey is a strong possibility. The only consolation is that they “might have prevented the war by publicizing the future events”. However, they predict that Turkish people residing in the US will be sent to concentration camps if this “inevitable war” erupts [36]. Not only are the authors themselves in awe of their masterpiece but it has been reported that the Turkish Foreign Ministry, alongside all ministers, the General Staff, and many young members of the Turkish army avidly devoured the book [37]. Be this as it may, the mystifying and revering attitude of the local press in conjunction with the intense self-promotion of the authors, portraying themselves as sages of the Oracle, led to a snowballing effect in marketing. The media, featuring the authors numerous times in news stories and TV programs, aimed

to join the bandwagon in profiting from the rising star of the authors. There was even an instance when a columnist from the mainstream press made the highly unlikely comment that American officials ‘whispered’ that there is highly classified information in the book [38]. Nonetheless, the book sold over half a million copies, which is considered a record in Turkey. Considering that it was originally circulated extensively before the publication, and there are many pirate copies in the market, the actual number of the buyers can be even more. More importantly, the book received very limited criticism except for Erciyes’ commentary.

Today, the two authors, who were struggling to make a living prior to the publication of *Metal Storm*, are bombarding the literary market in Turkey with successive sequels that they are now penning separately. One of the books portrays a war between Turkey and Israel, and according to its author Turna, it is bound to “develop antibodies among Turkish people who have treated Jewish people so good, and had nothing but betrayal in return” [39]. It is the Israelis that have stolen the body of Atatürk with the help of a “homosexual, Jewish convert called Rıfat Pamuk”. It is perhaps not coincidental that the last name of the ‘betrayal’ is Pamuk, after the famous novelist Orhan Pamuk who recently stood trial for stating that Armenian genocide took place. The mystery behind the Israeli plot is “cloning Atatürk and destabilizing Turkey” [39]. Of course, the book chants, there was never a Holocaust and Hitler was a puppet of the Jews.

The huge impact of *Metal Storm* in Turkey points to the fact that a part of the Turkish public sphere is stepping towards introversion and even a certain degree of schizophrenia regarding relations with the “West”. In that sense, *Metal Storm* was not a unique example of a popular culture product that preached the motto “Turks have no other friends, but Turks”; the TV series and blockbuster *Valley of the Wolves* also capitalized on the same theme.

5. Turkish Superhero as the Symbol of National Honor: Valley of the Wolves and Mafia Glamour

Valley of the Wolves was a hugely successful soap opera about an undercover Turkish agent who turns into an internationally revered mafia boss. The TV series was launched with the maxim, “This is a mafia serial” and lasted for three years, generating an immense fan group. When one of the main characters ‘died’ on screen, there were instances of funeral ceremonies, suicides, and even the beating of the actor who played the part of the murderer. The deceased character, Çakır, was a very ‘colorful’ mafia boss, who pillaged and killed, but somehow fitted into the ideal hero typology of the audiences. The key character of *Valley of the Wolves*, Polat Alemdar was an even more curious personality, who was said to be modeled on a real life mafia hit man and nationalist agent, Abdullah Çatlı.

Here a parenthesis must be opened for narrating a controversial chapter from Turkey’s recent history, related to the discovery of a clique that functions inside the state, conducting undercover right wing operations and maintains relations with the mafia. Since the 1970s there were rumors about state protected assassins who pulled the trigger in a number of unresolved murders of journalists, professors and engaged in covert sabotage operations against “Turkey’s enemies”. While their existence was never proven, a foremost journalist, Uğur Mumcu, who tried investigating the cases, was murdered. In 1996, a high ranking police chief, an alleged assassin sought after by Interpol and a parliamentarian were found to be traveling in the same car when they had a traffic accident nearby a small town called Susurluk. These murky relationships were never fully unearthed, but it is alleged that the former Turkish governments authorized the usage of state funds for the surreptitious activities of right wing hit squads, which were in return linked with mafia bosses and drug dealers [40]. The figures involved in this scheme were branded as members of the ‘deep state’ by the journalists reporting about the so-called Susurluk incident.

The web of entangled relations between state officials and mafiaesque agent/saboteurs remained unaccounted for as the cases were closed over time without answers being given. The success of *Valley of the Wolves* lay in its marketed image contending that its scenario carried insinuations to what happened in real life. In a way, the audiences were watching the series as if they were watching a documentary. Soon, fantasy began to get mixed up with reality. In real life, a jailed mafia boss,

Aladdin Çakıcı argued that he ordered the killing of the popular fictive ‘colleague’ of his, Çakır. He asserted this was because the character was based on him and did some acts he would never have done, like kissing the hand of another boss.

This ‘accomplishment’ of even attracting ratings by mafia bosses led to a total obsession with the series causing each episode to have top rating scores, even if their timing was “colliding with a national football match”. As a consequence of public uproar, after rumors that the series would end, the soap opera was turned into a blockbuster film called *Valley of the Wolves Iraq* in 2006. As can be deduced from the name, the film was about the epic deeds of the hero, Polat, this time in Iraq, but always dressed up in his elegant black suit. It must be mentioned that Polat, or rather in ‘real’ life Naci Şaşmaz is a shortish, dark man without the stereotypical looks of a jeune premier.

The film commences with the detention of the Turkish soldiers by the US infantry (again), and then it fantasizes that the Turkish officer in charge commits suicide with his “national pride broken”. The suicide note reaches secret agent Polat, who decides to take revenge on the ‘vicious’ Americans. Meanwhile, the US army is torturing Iraqis, killing innocent civilians mercilessly and a Jewish doctor is butchering the corpses of inmates in the tarnished prisons to ship their organs to rich patients in New York. The psychopathic US commander named Sam William Marshall (as in ‘Uncle Sam’ and ‘Marshall Aid’) is committing quasi-genocide, claiming that he is only executing God’s orders. The extremely simplistic storyline of the film becomes ‘sophisticated’ with the introduction of the idea that ‘crusade for democracy’ in the Middle East as uttered by the US President Bush, is in fact a crusade against Muslims. Overall, throughout the film there are a lot of nationalist propaganda motifs, alongside some mildly Islamic assertions regarding ethics that Muslims naturally own.

The release of the film was accompanied by so much fanfare and media support that all over Turkey crowds were cheering with happiness as the Americans were ‘beaten’ and the ‘justice’ was seen to done by the Turkish superhero Polat. The Prime Minister Tayyip Erdoğan and Turkish first lady had a special screening and reportedly remained ‘speechless’ with delight after the film [41]. Even a columnist renown for his sober views on mafia-state relations gleefully chanted how much he enjoyed the film, as it acted as an opiate to “sooth our bruised souls due to Western belittling and injustice” [42]. This cheering and clapping of the audience also spread to Turkish audiences across Europe.

On the whole, the public fascination grew and the film continued to break all records at the box office. The fact that semi-famous Hollywood actors starred in the movie version and even more famous stars paraded through the final episode of the TV series also contributed to the conviction that Turkey had achieved a global success with *Valley of the Wolves*.

In the interim, the creators of the *Valley of the Wolves* assumed the aura of mafia figures, becoming ‘untouchable’ by using bulletproof cars and bodyguards. The investigation regarding accusations that their production company used fake invoices to balance their financial transactions was stopped by the intervention of the members of the Grand National Assembly [43]. Meanwhile, a TV host/comedian was stabbed in his face because he called the main character of the *Valley of the Wolves* a ‘phony’ hero [44]. Most recently, the heroic hit man ‘Polat’ acted as a muse to the murderers of an Italian priest in the Black Sea coastal town of Trabzon and a judge of the High Administrative Court in Ankara [45]. The killer of the judge, who describes himself as ‘nationalist and Muslim’ was apparently involved in a plot organized by former army officers and ‘deep state’ figures, and he assumed the code name ‘Polat and the other confessed assassin also confirmed that he regarded ‘Polat’ as a hero. Still, these murder cases remain to be wholly resolved, regarding who else is involved other than the murderers in the plot, and if any such scheming exists.

Eventually, the producers decided to discontinue the series citing that they are blamed with the spree violence that burgeoned in secondary schools with the formation of youth gangs inspired by the hype of *Valley of Wolves*. Just as in the former examples scrutinized here, *Valley of Wolves Iraq* is perceived as recuperating national honor through asserting Turkey’s superiority. The motives of traditionalism, sacredness of national pride, superiority of Turkish culture, and the power of Turkey as a regional, and even world power are relished in both the TV series and their film version. Unlike the former two examples, *Valley of the Wolves* also praises conservatism and Islam; referred to as the two “permanent companions” of Turkish nationalism [46].

6. Conclusion

An analysis of selected popular culture works point out that certain themes are persistent in the contemporary Turkish national psyche. First of all, the invasion of Iraq and the detention of Turkish soldiers in Northern Iraq obviously caused dismay among the Turkish people. The instantaneous and unparallel success of the *Cola Turka* advertisement campaign right after the ‘sack’ incident proves that a lot of Turkish people sympathized with the campaign’s message of reverse cultural imperialism. As the subsequent commercial films of *Cola Turka* featured American soldiers lying down arms after drinking this soft drink, the creators of the campaign must be aware of the triangular relationship between the invasion of Iraq, the deteriorating relations with the US and rising nationalism in Turkey. In the cases of *Metal Storm* and *Valley of the Wolves Iraq*, the impact of the war in Iraq and rising anti-Americanism were even more straightforward as these two works specifically built on their themes on these very issues. Moreover, daily broadcast of unrelenting violence in the Middle East is giving the message to the Turkish people that their own security might be in danger. Especially now that the domestic tension with the Kurdish minority became manifest in terms of violence once more, security is a primary concern for the Turkish people.

At a time when Turkey’s internal and international political relations seem to be facing constant troubles, it is not difficult to see why all of the popular culture works studied here assert Turkey’s and the Turkish people’s pre-eminence. While the traditional alliance with the US is undergoing troublesome times, the relations with European Union are facing difficulties as well. The reform process towards EU *acquis* is stagnating and the frequent mentioning of the possibility of a “privileged partnership” for Turkey instead of full-membership status by some of the EU member states is strengthening the view that Europe does not want Turkey as a part of itself. Thus, it is no wonder that claims to Turkish superiority, such as Atatürk’s famous motto chanting that “one Turkish person equals to the whole world” find increasing resonance in Turkey. There is a clear ‘thirstiness’ for international recognition and approval, as the EU membership is perceived to be turning into an ever-elusive dream instead of an everyday reality.

References

- 1 Tarhanlı T., Kaya A. Türkiye’de Azınlıklar ve Anayasal Yurttaşlık (Minorities in Turkey and the Constitutional Citizenship). In: Tarhanlı T., Kaya A., editors. Istanbul: TESEV, 2005.
- 2 Öktem K. Return of the Turkish “state of exception”. Website of Middle East Report Online. Available at: URL: <http://www.merip.org/mero/mero060306.html>. Accessed at, June 4, 2006.
- 3 Cola Turka website. Available at: URL: <http://www.colaturka.com.tr/>. Accessed at, June 12, 2006.
- 4 Turna B., Uçar U. Metal Fırtına (Metal Storm). Istanbul: Timaş; 2004.
- 5 Kurtlar Vadisi (Valley of the Wolves) website. Available URL: <http://www.kanald.com.tr/dizi/kurtlarvadisii/>, Accessed at, June 12, 2006.
Kurtlar Vadisi Irak (Valley of the Wolves Iraq) website. Available at: URL: <http://www.kurtlarvadisiiirak.com/>, Accessed at, June 12, 2006.
- 6 Kasaba R. and Bozdoğan S. Turkey at a Crossroad. Journal of International Affairs. 2000; 54-1
- 7 Belgenet. Full list of legal harmonization packages with the European Union. Available from: URL: <http://www.belgenet.com/arsiv/ab/ab.html>. Accessed June 8, 2006.
- 8 BBC Turkish News Site. News story of about the protests held against scholars convening for a conference called Ottoman Armenians at the Last Phase of the Empire: Academic Responsibility and Problems of Democracy (İmparatorluğun Son Döneminde Osmanlı Ermenileri: Bilimsel Sorumluluk ve Demokrasi Sorunları). Available from URL: http://www.bbc.co.uk/turkish/europe/story/2005/09/050926_armenian_cnf.shtml. Accessed June 10, 2006.
Sabah Daily. News story about protests and debates regarding a conference called; Civilian Approaches to Problems of Democracy: the Kurdish Issue. Available from URL: <http://www.sabah.com.tr/2006/03/12/gnd101.html>. Accessed June 1, 2006.
Kadıoğlu A. Gurur, Korku ve İnanç (Pride, Fear and Belief). Radikal Daily. Available at: URL: http://www.radikal.com.tr/ek_haber.php?ek=r2&haberno=5911. Accessed at, June 10, 2006.

- 9 Çarkoğlu, A. The Rise of the New Generation Pro-Islamists in Turkey: The Justice and Development Party Phenomenon in the November 2002 Elections in Turkey. *South European Society & Politics*. 2003; 7.3: 123-156.
- 10 Turkish Daily News. News related to corruption in AKP. Available from URL: <http://www.turkishdailynews.com.tr/article.php?enewsid=45523>. Accessed June 2, 2006.
- 11 Özalp, Güven. New story about the 2004 European Union Report on Turkey's Accession Progress titled, AB İlerleme Raporu Tatlı Sert (EU Progress Report is both Sweet and Sour. Available from URL: <http://www.milliyet.com.tr/2005/11/05/siyaset/asiy.html>. Accessed June 5, 2006.
- 12 Hurriyet daily's internet site. News story on the refusal of the bill regarding Turkey's military involvement in the invasion of Iraq. Available from URL: <http://www.hurriyetim.com.tr/haber/0,,sid~1@w~342@tarih~2003-03-01-m@nvid~237333,00.asp>. Accessed June 5, 2006.
Hurriyet Daily website. Tezkere reddedildi (The bill is rejected) Available from URL <http://arsiv3.hurriyet.com.tr/anasayfa/0,,tarih~2003-03-01-m,00.asp> Accessed June 6, 2006.
- 13 Website of the Prime Ministry's Media, Publishing and Information General Directorate of the Turkish state. International news coverage of the reactions concerning the rejection of the bill on Turkey's military involvement in the invasion of Iraq. Available from URL: <http://www.byegm.gov.tr/YAYINLARIMIZ/bultenler/tezkere/arsiv/tezkere-2.htm>. Accessed June 5, 2006.
- 14 Hurriyet Daily website. Once Siir Sonra Tezkere (First the poem than the motion). Available from URL: <http://www.hurriyetim.com.tr/haber/0,,sid~1@w~1@tarih~2003-03-21-m@nvid~245323,00.asp> Accessed June 6, 2006.
- 15 Çarkoğlu A., Kirişçi K., Ersan R., Yılmaz H. Public Opinion on the European Union Membership in Turkey. Website of TESEV – (Türkiye Ekonomik ve Sosyal Etüdler Vakfı) Turkey Economic and Social Studies Foundation). <http://www.tesev.org.tr/temmuz2002/t1.html>. Accessed May 28, 2006.
Çarkoğlu A. Who wants full membership?: Characteristics of Turkish public support for EU membership. In Çarkoğlu, A, Rubin, B. Turkey and European Union: Domestic politics, economic integration and international dynamics. London: Frank Cass Publishers; 2003.
- 16 Yılmaz H. Euro-skepticism in Turkey. Infakto Public Opinion Surveyor Company's website. Available at: URL: <http://www.euractiv.com/ndbtext/turkey/survey.ppt#6>. Accessed June 30, 2006.
- 17 Özkırmı U., Uyan P. Türkiye'de Milliyetçilik Araştırması 2006. (Survey on Nationalism in Turkey 2006). Radikal Dailly website, interview with Dr. Özkırmı. Available at: URL: <http://www.radikal.com.tr/haber.php?haberno=187334>. Accessed June 10, 2006.
- 18 Kirişçi, K. Turkey. In: Starvidis, S. Couloumbis, T., Veremis T., Waites, N., editors. The foreign policies of the European Union's Mediterranean states and applicant countries in the 1990s, London: Macmillan Press. 1999.
- 19 Kadioğlu, A. Kendini Beğenmeyen Gururlular (Proud and Self-hating). Radikal Daily Website. Available at: URL: http://www.radikal.com.tr/ek_haber.php?ek=r2&haberno=5757. Accessed at , June 11, 2006.
- 20 Billig, M. Banal Nationalism. London: Sage; 1995.
- 21 Özkırmı U., Yumul A. Reproducing the nation: 'Banal nationalism' in the Turkish press. *Media, Culture & Society*. 2000. SAGE Publications Vol. 22: 787–804.
- 22 Bali R. Yeni aristokratlar: köşe yazarları (New aristocrats: the columnists). *Birikim*. 117: 48-56; 1999.
- 23 Bali R. Ertuğrul Özkök: Yeni Türk insanın heykeltraşı (Ertuğrul Özkök: Sculptor of the new Turkish people). 144: 57-65; 2001.
- 24 Özkan D, Foster R.J. Consumer citizenship, nationalism, and neoliberal globalization in Turkey: the advertising launch of Cola Turka. *Advertising and Society Review*. Vol: 6(3); 2005.
- 25 Star Daily. Surge in sales of new Turkish cola seen as backlash against US. Available at: URL: <http://www.thestar.co.za/index.php?fArticleId=192530>. Accessed at, June 12, 2006.
The Globe and Mail website. Chevy Chase stars in Turkish Commercial. Available at: URL: http://www.theglobeandmail.com/servlet/Page/document/v4/sub/MarketingPage?user_URL=http://

- /www.theglobeandmail.com/%2Fservlet/%2Fstory/%2FRTGAM.20030718.wchase0718/%2FBNStory/%2FEntertainment/%2F&ord=1150348182952&brand=theglobeandmail&force_login=true. Accessed on, June 2, 2006.
- 26 Tempo Weekly. Cola-yi Milliye. 29, July 17-23; 2003. Available at: URL: <http://www.tempoergisi.com.tr/dergi/02711/index.php>. Accessed at, June 12, 2006.
Strategies for harnessing local brands. Indian School of Business Magazine. Available at: URL: http://www.isb.edu/isbinsight/ISBInsight_March2006.pdf.
 - 27 Internet site of a popular comic site. "Drink it or leave it!" (Ya iç , ya terket). Available at: URL: <http://www.gag.web.tr/yazi.php?yad=843>. Accessed at, June 1, 2006.
 - 28 Kalkan, N. Milliyetçi' reklamlar alafranga çocuklara tepki (Nationalist advertisements are reactions against a la Franga brats). Tempo weekly. 957: April 2006.
Alpman Nazım. We are not a la Franca brats. Aksam Daily website. Available at: URL: <http://www.aksam.com.tr/arsiv/aksam/2003/07/14/yasam/yasam2.html>: Accessed at: June 11, 2006.
 - 29 Milliyet Daily Website. Interview with Ali Taran. Available at: URL: <http://www.milliyet.com/2004/10/07/business/bus06.html>. Accessed at, June 1, 2006.
 - 30 Bora, T. Cem Uzan ve Neo-faşizmin yeni yüzü: şirket konserlerinden partiye (Cem Uzan and the new face of neo-fascism: from corporate concerts to political party). Birikim:162:53-60, 2002.
 - 31 Ulker website. Available at: URL: <http://www.ulker.com.tr/>. Accessed at, June 12, 2006.
 - 32 Widmer, T. Death to crusade. New York Times website. Available at: URL: <http://www.nytimes.com/2005/09/18/books/review/18widmer.html?ex=1150344000&en=32ea537e8aaf45f0&ei=5070>. Accessed at, June 1, 2006.
 - 33 Helvacı C., Alonso R.N. Borate deposits of Turkey and Argentina; a summary and geological comparison. Turkish Journal of Earth Sciences. Vol. 9: 1-27; 2000.
 - 34 Erciyes C. Fırtına'nın Alçak Basıncı (Low Pressure of the Storm). Radikal Daily Book Review Edition. April 1; 2005.
 - 35 Website of Timaş Publishing house.
Available at: URL: <http://www.timas.com.tr/kitap.php?id=863>. Accessed at, June 12, 2006.
 - 36 Vatan Daily. Önce İstanbul ve Ankara'yı vuracaklar (First they are to bomb Istanbul and Ankara). 2004, 19 Dec.
 - 37 Radikal Daily. Dışişleri Bakanlığında Metal Fırtına (Metal Storm at the Foreign Ministry). 2005: 17 Jan.
Sabah Daily. Kabine Metal Fırtına'yı okuyor (Cabinet reads the Metal Storm). 2005: 16 Jan.
Vatan Daily. Metal Fırtına 2 ve Derin İmparatorluk Geliyor (Metal Storm 2 and Deep Empire is on the way). 2005: 17 Jul.
 - 38 Ozkan F. ABD'yi Uc Hassas Nokta Kızdırmış (Three sensitive issues made the US angry) Available at: URL: <http://www.radikal.com.tr/haber.php?haberno=148351> Accessed at, June 12, 2006.
 - 39 Hepkon H. 40 katır mı, 40 satır mı? (40 mules or 40 knives). Radikal Daily Book Review Edition. 2005: 23 Oct.
 - 40 Bilkent University Political Science Department website. Grand National Assembly's report on the Susurluk incident. Available at: URL: <http://siyaset.bilkent.edu.tr/susurluk/tbmm/>. Accessed on June 1, 2005.
Er, A. En 'bağımsız' derin devlet (Freest of all, the deep state). Commentary website called 'Gazetem'. Available at: URL: <http://www.gazetem.net/aeryazi.asp?yaziid=213>. Accessed at, June 1, 2006.
 - 41 Turkish Daily News website. Sheep in wolves' clothing.
Available at: URL: <http://www.turkishdailynews.com.tr/article.php?enewsid=36465>. Accessed on, June 5, 2006.
Haber 7 news website. Emine Erdoğan'dan Kurtlar İtirafı (Wolves' confession from Emine Erdogan). 2006; 4 Feb.
Available at: URL: http://www.haber7.com/haber.php?haber_id=136289. Accessed on, June 12, 2006.

- 42 Dündar, C. Saflara hoşgeldin Polat! (Welcome to our side Polat). Milliyet Daily website. Available at: URL: <http://www.milliyet.com.tr/2006/02/05/yazar/dundar.html>. Accessed on, June 12, 2006.
- 43 Milliyet Daily website. Kurtlar Vadisi’de izlendi (Valley of the Wolves was also taped), Available at: URL: <http://www.milliyet.com.tr/2005/05/19/siyaset/asiy.html>. Accessed on, June 11, 2006.
- 44 Hurriyet Daily website. Metin Uca bıçaklandı (Metin Uca is stabbed). Available at: URL: <http://www.hurriyet.com.tr/sondakika/4405275.asp>. Accessed on June 12, 2006.
- 45 Zaman Daily website. Saldırganın kod adı Polat (Code name of the assailant is Polat). Available at: URL: <http://www.zaman.com.tr/?bl=haberler&alt=&trh=20060525&hn=288184>. Accessed on June 11, 2006.
- 46 Bora T. Türk Sağının Uc Hilali: Milliyetçilik Muhafazakarlık, İslamcılık (Three Crescents of the Turkish Right: Nationalism, Conservatism and Islamism) Istanbul: Birikim; 1998.

In Search of the Invisible Balkan Cinema

Efthimios Hatzis¹, Erica Sheen², Sean Homer³

¹SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, ehatzis@seerc.info

²The University of Sheffield, Dept. of English Literature, Western Bank, Sheffield, S10 2TN, UK, e.sheen@sheffield.ac.uk

³City College, Dept. of Business Administration and Economics, 2 Kalapothaki Str, 54624, Thessaloniki, Greece, s.homer@city.academic.gr

The invisible in cinema is defined as the ability to facilitate audience participation with the art form in a manner beyond story, events, locales and characters. The invisible provokes a new meaning that indicates perhaps towards the “divine”, or simply the essence of human relationships. This is clearly a director’s intention of how she uses her cinematic tools for constructing the cinematic project. The notion of descriptive and expressionistic approaches in film directing are discussed to indicate the cinematic scheme directors can utilize to express the subtext of the screenplay. The expressionistic approach to directing is prone to Auteur directors and the invisible as well. An example of a cinematic scheme utilizing sound design with the image in provoking the invisible is discussed from the Balkan director Angelopoulos. Citations from cinema studies and other visual arts are discussed to build the framework for a concept of the invisible. This will serve as the point of departure for further analysis of Balkan cinema and its cultural and geographical correlations. The analysis of producer’s intent and market analysis is utilized to shed light on the production difference between Balkan and mainstream cinema.

Keywords

Auteur, cinema, dramatic analysis, invisible, sound design.

1. Introduction

The question posed in this paper is what is invisible cinema and further what cultural and geographical correlation may this have in Balkan Cinema. First I will define what I understand as invisible cinema. Second, I will show how a film director, as the author of this art form, utilizes his/her craft to imply the invisible. Finally, this will serve as the starting point or prism to observe and gather cultural and geographical correlations and comparisons for films produced in the Balkan region, with respect to world cinema.

2. Approaches in fiction film directing

In fiction cinema perhaps a more effective form of communication with an audience is not only the story but also the way the story is told. As a raw illustration: one person can execute the telling of a joke in a splendid manner and produce the desired result, uncontrollable laughter. While another person may execute the same exact joke with a minimal result. Thus it is not only the actual “text” of a joke, but also the manner in which a joke is told that also triggers a result in the audience. This analogy holds true in filmmaking as well. It is not only the story or screenplay that induces a reaction in the audience, but also the manner in which this story is presented. The executer of the joke in this case is the film director.

The question arises what is the role of the director? Does s/he simply act out and record the screenplay? Or does s/he create a new art form based his/her interpretation of the screenplay? Robert Bresson the French filmmaker had similar thoughts on this matter, to quote Bresson:

Two types of film: those that employ the resources of the theatre (actors, direction, etc.) and use the camera in order to *reproduce*; those that employ the resources of cinematography and use the camera to *create*. (p: 5, [1]).

Bresson points to two types of film production here: “reproduction” and “creation”, and this has to do with the manner a story is told. My supposition is that in order “to reproduce” the film director need not employ his/her cinematic tools (such as camera, art direction, lighting, acting style, etc.) to *express* the essence and theme of the story. While in order to “create” directors do use the cinematic means to express their own interpretation of the subtext of the screenplay. In other words, the director “creates” a new and different art product from the screenplay. Therefore, I consider the “reproductive” approach to directing as *descriptive* and the “creative” approach as *expressionistic* (Fig.1).

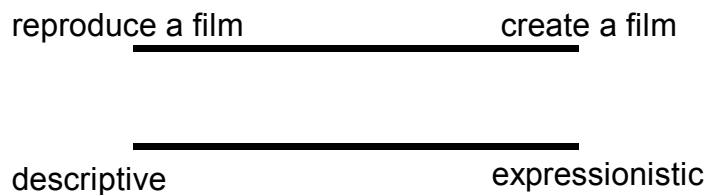


Figure 1 Polar ends of directing

These two ends of the pole - descriptive and expressionistic – can be one facet in the gamma of the film director’s role.

2.1 Descriptive and expressionistic style in directing

The descriptive end of the pole is more indicative of television studio series and *multi camera* direction. Here the director’s responsibility is to record, or “describe” the events and actions of a scene. The set up of two and more cameras in television studio direction lends to a non-expressionistic approach, in the sense that the camera position is chosen due to practical considerations such as a limited three wall stage, and placing cameras as to not be obstructing each others view. Notice the camera placements on the overhead schematic are positioned as not to be in the angle of view of each other, and the limited possibility of camera placement due to the lack of a fourth wall (Fig.2).

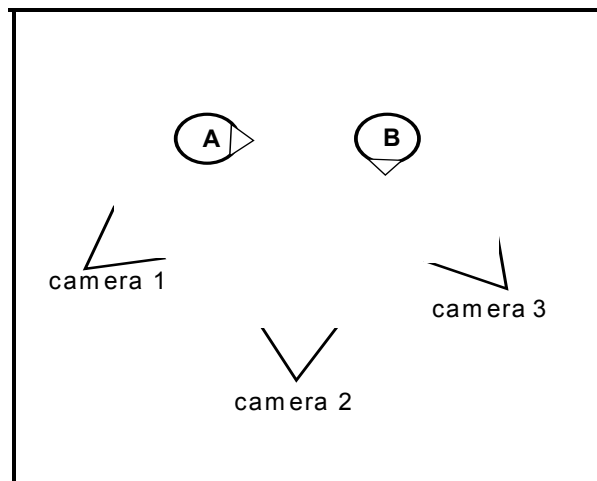


Figure 2 Example of television camera positions, scene with two actors (A, B)

Camera positions in a fiction film are more “expressive” because they are not limited to practical boundaries, along with the parameter that a film director utilizes *one camera* to cover a scene. The use of only one camera is the norm of shooting, regardless of the scope of the production and budget. This holds true for the entire gamma of film production from Hollywood studio productions to independent

Balkan productions. Thus if a director chooses to cover a scene in four different camera angle or shots, the scene will be repeated in its entirety four different times in exactly the same manner, and the camera will move and be positioned in a different “set-up”. This is different from television series coverage where the multi camera set up all cameras record simultaneously.

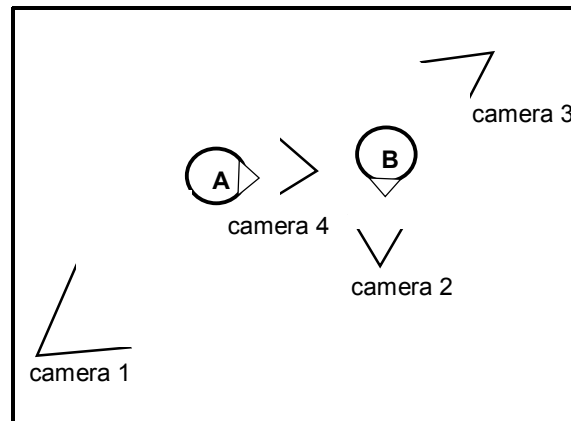


Figure 3 Example of cinema camera positions, scene with two actors (A, B)

Notice in the overhead schematics Fig. 3: that camera 1 and camera 3 would have other cameras in their line of vision if shot simultaneously. Note as well, that an art directed “dressed” fourth camera lends more possibilities to the placement of camera 3.

This manner of working in cinema hides the primary difference between television direction of series and cinema direction. Film directors actually prefer to work with one camera because the one camera is more expressive allowing the ability to construct every shot with attention to detail. In this manner the director can work on the detail of every shot: The visual elements: lighting, the art direction, the composition of the frame. And the acting range as well (an actor must perform differently on a long shot than that of an extreme close up). The purpose of every shot can be to communicate with the audience on more than one level.

2.3 Macro Dramatic Analysis.

Using only one camera during shooting aides an expressionistic style due to the freedom of camera placement and the ability to pay attention to detail, but it is not the only element of expressionistic directing style.

Another element of the expressionistic approach in film direction deals with the dramatic analysis of the screenplay and the theme. If the same exact screenplay was given to five directors with the same actors, locations, and art direction five different films will be produced. This has to do with the director’s interpretation of the screenplay.

A film director devises cinematic schemes in order to express and highlight the macro dramatic analysis. The macro dramatic analysis is based on the turning point the director believes a screenplay has in order to express the theme of the film. Screenplays are constructed so that they usually have three major turning points. Syd Fields classic work, *Screenplay* underlines the structure of a screenplay based on these three turning points. Based on these turning points the director outlines a narrative progression breaking down the screenplay in units. Then the director can devise a *cinematic scheme* utilizing his/her directing tools to underline the theme or *subtext* of each unit in the dramatic progression. The Soviet filmmaker and instructor, Sergei Eisenstein, refers to this role of the director:

S.M. (Eisenstein) looked on the work of the film director as a direct continuation of the creative work that begins with the scenario writer, and consequently he (Eisenstein) attached special importance to the director’s ability to find the *production solution* that should express the ideas of the scenario most exactly and

vividly from every view-point including that of the compositional structure (p: 12, [3])

Here Eisenstein mentions *production solutions*, in other words using the form, and technical aspect of directing a movie, or constructing *cinematic schemes* to express not only the story but also the essence or theme of the story.

2.3.1 Example of Cinematic Scheme Based on Dramatic Analysis.

An example of cinematic schemes can be readily found in Sidney Lumet's Twelve Angry Men. Based on his book Making Movies [4] and his Master's class[5] I attended in New York, I will outline the cinematic schemes of this film. This is good example because of the limitation that the entire film takes place in one room.

In the first act: a teenage African-American is on trial for murdering his father. The 12 Jurors are then enclosed in the meeting room to come to a verdict. All seems very easy at first all these men will vote guilty, until the first turning point where one man votes to consider the case in more detail. Lumet uses a very specific cinematic scheme that follows the progression of the story. For this first part of the film he utilizes the coverage of the film in the following manner:

- The use of a *wide-angle lens*, which makes the room seem larger and the spaces between the characters farther apart.
- A camera height placement was *above eye level*. Giving the feeling that the jury is not meeting eye to eye.
- The grouping of characters in every shot. No character has been seen alone in a close up in one shot.

As the story moves forward the cinematic scheme progresses as well. In the second act the jury start to discuss the case, and find there are many holes in the case. All start to become convinced of his innocence, except for one man who insists the boy should be punished. This is the second turning point. Here the cinematic scheme changed:

- The camera had a *normal lens*. This made the space appear smaller, and the characters closer together.
- The camera height was at *eye level*. This aided the sense of the characters, as well as the spectators, seeing "eye to eye."
- There was less grouping of characters in a shot. The character that first entertained the idea the boy not being guilty was the very first close-up of a juror. As the second act progressed, more and more close-ups were used as people became convinced of his innocence.

In the final act the one man who insisted on a guilty verdict became intense and personally abusive of the other members. In the climax of the film his rage leads to the revelation that he was making a decision based on chauvinism. Of course the camera scheme highlighted this as well

- The camera was mounted with a *long (tele) lens*. A long lens has the characteristic of compressing space. This gave the illusion that the room is small and increasing the immediacy of the interpersonal proximities of the scene.
- The camera's height was *below eye line level*, thus giving more dramatic intensity, and the ceiling of the room was in view. This coupled with the long lens created the sense of claustrophobia. As if the room was steadily getting smaller and smaller.
- Most of the shots were single close ups of the characters. Thus the men now are not a mass as filmed in the first act, but each man has a persona, hinting that the boy on trial has a persona as well, and is not only a bias stereotypical figure.

In the resolution of the film where the verdict was agreed, the *camera level was extremely high* and with a *wide-angle lens* allowing, as Lumet says to give us air, and to breathe.

This cinematic scheme Lumet coins “lens plot” (p:81, [4]), in other words the camera will tell us the story as well. One can have the volume turned down and the essence of the story will still be communicated. Lumet definitely used the lens and camera to express the theme, that men need to analyse and see each other as a person, face to face, to find the truth. The visuals resonate this theme. This approach to film directing is placed towards the expressionistic polar end.

The lens, and camera height level is not the only element that can be utilized expressionistically. All choices a director needs to make can express the theme of the film from the colour of the walls, composition of a frame, to the acting style and the music.

2.4 Auteur Filmmakers

My premise is that the invisible in film can be found on the expressionistic end of directing approaches. Further more directors who are on the expressionistic pole are usually what are commonly called: auteur, or authors of a film. Directors who create a new world, a new gestalt and its inner logic, that is different from the creation of the screenplay. Auteur directors have a recognizable signature, and a distinctive personal expression in their cumulative body of work. Further more, auteur film directors generally produce independent and non-mainstream films. Therefore, non-mainstream, auteur films are more prone to the invisible.

The Balkan region is filled with great Auteur films with international acclaim, a partial list would include: Emir Kusturica, Theo Angelopoulos, Yilmaz Guney, Dusan Makavejev, Lordan Zafranovic, Milcho Manchevski, Zelimir Zilnik, Borislav Sajtinac, Omer Kavur, Stavros Tornes, Zeki Demirkubuz, Lucian Pintilce, Fracios Lunel, Eduard Zahariev and Krassimir Kroumov.

Balkan production lends itself to auteur filmmaking based on producer’s intent. A Balkan film can rarely be a mainstream movie due to two factors: budget and language. Lets take a look at mainstream movies in order to have a perspective on Balkan productions.

3. Mainstream Movies

Mainstream movies usually have a highly structured narration and rhythm of turning points, a flow of positive and negative events in scenes, with the goal to hook the audience through out the duration of the movie. This predominates in Hollywood produced films, with the illusion of engaging the audience and keeping them in total suspense. There is no lagging, or moments of boredom for the audience. If the storyline is blueprinted well enough via the classical formulas that many how-to-write-a-screenplay authors¹ outline, the movie will achieve its goal, to entertain the audience through out the film with out having their mind wander.

However, not allowing our mind to wander does not allow the spectator to be involved or to participate with the film on more levels, film theorist Andrew Horton calls this a *cinema of contemplation*. There is something that I cannot define yet. Perhaps it has to do with the pace of the story, and the editing, the speed of changing images is phonologically quite enthralling -- or is it really the opposite? Is it possible that this rapid speed of images causes our brain to go into an alpha state? Comparable to a mass induced biofeedback that synchronizes our brain waves to the state before sleep. An interesting study would be using an electroencephalogram (EEG) to record the wavelength of the audience.

Rhythm and editing pace in film is an area for interesting research. Horton in analyzing the renowned Greek auteur, Theo Angelopoulos, elaborates on his signature long takes, and slow rhythm as an element to contemplation.

Angelopoulos’s extremely long takes, often lasting up to ten minutes, suggest the “continuous” relationship between Byzantine art and the observer: one can stand and gaze at an icon, a program, or a scene as long as one wished. In this sense it

¹ Syd Fields, Robert Mckee, Linda Cowgill, Christopher Vogler, and Denny Martin Flinn are a few examples.

was the observer who controlled how long his or her experience would last. While a viewer in cinema is locked into the forward flow of images, the lack of classical editing and traditional narrative pacing seen in Angelopoulos's work means his films are unique in allowing the audiences time to "look around" within each scene as they please. Such a lack of narrative drive results in the establishment of a more personal and contemplative relationship between the film and the viewer, much like that between the observer and the icons in an Orthodox setting. (p. 30, [6])

Horton raises a crucial issue here that could be a key difference between films created by Auteur and mainstream films, the audience's participation, or identification.

3.1 Demographics of Theatre Going Audience

The intent of a film can be determined by the reason it has been produced, for a producer this may be market analysis.

Based on the demographic analysis in the trade paper, "Variety", presented in the European producer's training seminar [7], the target groups of theatre going audience is the same internationally, and is broken down in two major categories. The analyses of the demographics are as followed: The first target group is in the age group of 18-24 year olds: young males (YM), and young females (YF). This is the group that attends movie theatres on average 1.4 movies per week.

The second group is 18-24 year olds [same as above?]: old male (OM) and old female (OF). This group attends the movie theatre on a significantly lower average of 7 movies a year (Fig 4).

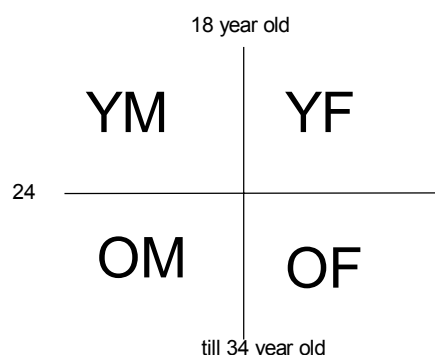


Figure 4 Market analysis of theatre going audience

Mainstream movies are made exclusively for the YM and YF target group, the rate of theatre attendance in this group justifies the large budget of a mainstream production. Producers preparing an independent or less commercial project have the older group, OM and OF targeted as an audience. Thus their budgets need to be significantly lower to justify their returns.

3.1.1 Theatrical Distribution as Advertisement

Theatrical distribution is important to producers because it will boost the financial livelihood of the product after its theatrical release. As pointed out in Gregory Goodell's [8] text receiving even minimal theatrical distribution will have significant positive correlation in video and DVD, cable, direct to television, and other sales venues. Thus the theatrical distribution is crucial for a producer and acts as solid long-term advertisement. Producers may accept a loss in the theatrical distribution of a picture in order to secure good payback in the post-theatre venue. In my interview with Hollywood director and producer Paul Aaron (September 2005), Mr. Aaron mentioned cases where a producer is willing to actually absorb the costs of renting several movie theatres in major US cities, self-distribution, for the livelihood of a movie in post-theatrical sales. Thus the demographics of theatre going audience are crucial in the producer's long-term investment.

3.1.2 Mainstream Producer's Intent

The budget for a theatrical release advertisement for a mainstream movie is 50%-100% of the budget of production of the movie. Commercial movie budgets on average range between \$40,000,000 and \$150,000,000. The mainstream producer is concerned if this money will make its returns, and the advertisement will be effective. In further analyses by "Variety": the mainstream target group YM (young males) is the most responsive to advertisements, it has been shown that they are highly influenced by ads, and their decision period to buy is -- seconds. One advertisement, one trailer, a poster will hook them into a purchase - attending the theatre.

Interestingly, or one might say tragically, enough in the analysis of the demographics it has been shown that the group YF (young females) follow the decisions that the young males set. The YF group does not have another assessment pattern, only the trend set by YM. Thus mainstream advertisements are targeted at young males 18-24 year old exclusively. Therefore a mainstream producer selects a screenplay, develops it, and packages the film to appeal to the young males. This should not be a surprise, if we walk through our local video store the images that pop from the aisles emphasizes this point. Perhaps this supports the notion of the *male gaze* in Laura Mulvey's essay Visual Pleasure and Narrative Cinema [9]. Where the narrative, the camera, and the spectator is through a male's point of view. The *intent* of the producer is to attract this target group. On the other hand, it is crude to generalize and say all mainstream movies are produced with this intent.

The intent of communication is crucial to me as a filmmaker; I find the intent to be a more powerful form of communication than the narrative. This is an area of interest for more research.

3.1.3 Art films and Balkan productions

The older target group OM and OF do have a high level of response to advertisement, they are influenced by critical reviews, word-of-mouth, and festival awards. They are more concerned about the time they are investing in a movie theatre to be worthwhile, the product be of quality. The films this older group tend to watch is over-generally coined *art film*. These are movies that are not screened in multiplexes, but art house cinemas, and festivals. A producer making a film for this target group is more likely to invest in an Auteur filmmaker of international status. The production budget will be a fraction of a mainstream movie and probably independently of with government subsidised support. The invisible is prone to appear in this category of production due to production intent and the auteur filmmaker.

Balkan film production usually falls into this category, and a clever producer will bank on the international success of a film in this small target group of the art house film, though in their own country they may prove not to be a financial success. Such is the fate of Angelopoulos in Greece. Here the intent of the film can be cultural, historical, political and personal.

Another form of Balkan productions is a film made exclusively for the YM and YF target group of the country of production. The producer does not have a goal to receive international distribution, films like the Greek box office "Safe Sex" and "Frozen Stiff" in Yugoslavia. In an interview [10] Milorad Milinkovic, the director of "Frozen Stiff" noted that the financing of his film came from private sources that expected to make their returns in Yugoslavia alone.

4. In Search of the Invisible

4.1 Schrader's Transcendental Style in Film

Invisible, but film is a visual medium? I'm not sure how to define it just yet. Is it a subcategory of expressionistic approach of directing, and is prone to auteur filmmaker? Perhaps the writer and director Paul Schrader touches upon it in his work "Transcendental Style in Film

In recent years film has developed a transcendental style, a style which has been used by various artists in diverse cultures to express the Holy. Just as anthropologists at the turn of the century discovered that artisans in unrelated cultures had found

similar ways to express similar spiritual emotions, so, in cinema, unrelated filmmakers have created a consensus of transcendental style. The style is not intrinsically transcendental or religious, but represents a way (a Tao, in the broadest sense of the term) to approach the Transcendent. The matter being transcended is different in each case, but the goal and method are, at their purest, the same. (p 3, [11])

In his publication Schrader analysis three auteur filmmakers from different cultural and religious backgrounds: Yasujiro Ozu in Japan, Robert Bresson in France, and Carl Dreyer in Denmark. He takes a film analysis and visual analysis approach to defend his premise. He studies camera work, editing, camera placement, visual composition and casting to exemplify the transcendental style.

4.2 Barthes's' *Studium and Punctum*

In his meditation on the photograph, 'Camera Lucida', Roland Barthes explores the notion of the image communicating in a two-fold manner. One is the *studium*: the average cultural, historical, political meanings a spectator receives from the photographs. The other is the more private and personal the *punctum*:

it is this element which rises from the scene, shoots out of it like an arrow and pierces me. A Latin word exists to designate this wound, this prick, this mark made by a pointed instrument: the word suits me all the better in that it also refers to the notion of punctuation, and because the photographs I am speaking of are in effect punctuated, sometimes even speckled with these sensitive points; precisely, these marks these wounds are so many points. This second element which will disturb the *studium* I shall therefore call *punctum* (p 27. [12])

The *punctum* is where the invisible is, beyond historical, political, cultural groundings. The *punctum* is a place as Barthes mentions "bruises me" it moves or creates tension on a highly personal level beyond the *studium* --the specifics of time, place or events.

4.3 Tarkovsky on the Poetry of Cinema

Cinema is a new art form, and can communicate on many levels beyond story and this is where the invisible exists. Andrei Tarkovsky, the Soviet film director has created a school in poetic cinema. He mentions the possibility of communicating in another manner to a traditional linear storyline, which lends to a more active participation of the spectator.

But film material can be joined together in another way, which works above all to lay open the logic of a person's thought. This is the rationale that will dictate the sequence, and the editing which forms them into a whole. The birth and development of thought are subject to laws of their own, and sometimes demand forms of expression which are quite different from the patterns of logical speculation. In my view poetic reasoning is closer to the laws by which thought develops, and thus to life itself, than is the logic of traditional drama. And yet it is the methods of classical drama which have been regarded as the only models, and which for years have defined the form in which dramatic conflict is expressed.

Through poetic connections feeling is heightened and the spectator is made more active. He becomes a participant in the process of discovering life, unsupported by the ready-made deductions of plot or ineluctable pointers by the author. He has at his disposal only what helps him penetrate to the deeper meaning of the complex phenomena represented in front of him. (p 20, [13])

Tarkovsky is expressing the concept of the *punctum* in film. Tarkovsky's "poetic reasoning" and "poetic connections" is a participation that requires this approach by the author of the film, the director, and the spectator as well.

4.4 Cinematic Tools: Example in Sound Design on Angelopoulos' Eternity and a Day

The director in order to communicate with the audience on the level described here uses his cinematic tools expressively, he creates cinematic schemes that lead to the essence of the film, or the subtext of the narrative.

One cinematic tool a director has is the sound design. Theo Angelopoulos in his Cannes winner 'Eternity and a Day' utilizes the sound design in a specific scene that I would like to analyse.

The scene is 'the bus scene' where the dying writer and an Albanian boy he has saved from child exporting [do you mean exploitation?] get on a bus as a whimsical [pass time?], not to go somewhere, but as a last ride before they depart for different countries –the boy to his home in Albania and the writer to hospital for the terminally ill in Italy. The sequence opens and closes with the same location a bus stop on the pier in Thessalonica. The bookend of the same stops gives the scene closure, a full cycle in time. But lets talk exclusively about the sound design in this sequence. When the old fashion bus appears we hear:

- streaking of the breaks
- old fashion knocking of the bus engine
- footsteps,
- rain
- passing cars
- pier ambience
- night city ambience.

All these sounds are heard realistically.

Once we are in the bus slowly the sound design changes to an expressive style. The ambience of the city and the pier fade to silent. The breaking of the bus and the old fashion purr of the bus engine are no longer heard Bicycles passing outside are silent. All the sounds of the outside world are silent. The opening of the bus door and windows are heard at full volume. The footsteps in the bus are exaggerated with loud squeaking, resembling to the sounds actors make on a wooden plank stage. Other sounds are exaggerated as well such as the brushing of clothes and the shifting of weight. The dialogue is heard pristine clear, with no additional ambience as if in a small room. As if the bus is a theatrical stage no outside sounds are heard and the bus itself has turned into a sound vacuum no ambience is heard (the purr of the engine, the movement the breaking etc.)

Thus the exaggerated Foley sound (sounds actors make from their movements brushing clothes, footsteps and such) and the absence of all other sounds allow for this scene to transcend. This scene is not just descriptive: a whimsical bus ride that Alexander and the boy are on, where people get on and off and then the ride is over. With the use of this sound design, the scene becomes expressive: the bus ride becomes a microcosm of time, the desire for the boy and Alexander to be together surpasses the limitations of time they have, and the bus ride becomes a metaphor of life. You travel with someone in life as other come and go in your life till the end of the road, where death lies. Here again Angelopoulos plays with time by placing a Greek poet from the previous century, Solomos on the bus! Thus, even the time of death is eternal and not linear. This scene transcends in order to express the theme of the entire film as the title shares: Eternity and a Day, or if you wish, eternity and a bus ride.

In a personal interview with the sound mixer of 'Eternity and a Day', Constantine Varibiotis, we were told that the scene was designed to have all the realistic sounds on the bus as well (including the bus engine, the city ambience etc.) But the scene did not work, it did not transcend. He worked with Angelopoulos by removing sounds and exaggerating others. The same images of the scene came to life only after this expressionistic approach to sound design. The punctum here resides in the juxtaposition of the sound design and the image.

4.5 Why invisible?

I am still searching for why the word invisible works for me as opposed to poetic or transcendental or expressive, the invisible is all of these and something more as well. What stands out for me in cinema are moments, sometime glimpses, something that pricks us or wounds us as Barthes calls it, a place where time is suspended. Russian Philosopher and theologian Pavel Florensky considers the divine implications of this condition:

In the apostolic Creed, God is named as “Maker of all things visible and invisible.” These two worlds –the visible and the invisible –are intimately connected, but their reciprocal differences are so immense that the inescapable question arises: what is their boundary? Their boundary separates them; yet, simultaneously, it joins them. How do we understand this boundary?

Here, as in any difficult metaphysical question, the best starting point always is what we already know in ourselves. The life of our own psyche, yes our own soul’s life, is the truest basis upon which we may learn about this boundary between the two worlds. For within ourselves, life in the visible world alternates with life in the invisible, and thus we experience moments –sometimes brief, sometimes extraordinarily fleeting, sometimes even the tiniest atom of time-when the two worlds grow so very near in us that we can see their intimate touching. At such fleeting moments in us, the veil of visibility is torn apart, and through that tear –that break we are still conscious of that moment –we can sense that the invisible world (still unearthly, still invisible) is breathing: and that both this world and another world are dissolving into each other. (p.33, [14])

The punctum for me are moments in cinema where this boundary between the visible and the invisible are alternating even for a moment, a glimpse, this is what interests me as a filmmaker as well.

4.6 Why Invisible in Balkan Cinema?

The worldwide view of Balkan Cinema maybe what Dina Iordonavich coins the *Cinema of Flames* [15].

In a conference at Yale University on Balkan Cinema the following question was posed:

Today the very word “Balkan” still tolls with a deeply ominous tone. Situated on the fringe, indeed at the bottom of the West, they have often been characterized as uncivilized, anarchic, abject. Despite venerable traditions of literature and popular art, Western universities have paid the region scant respect. But a robust cinema has brought this region and its politics forcefully to our attention. Ravaged by conflict and dangerously minded, this “Power keg” of Europe has served as a charged movie set for local productions to exploit its spectacularly varied topography and its diverse language, religions, and folk rituals.

There are other cultural elements not linked exclusively with historical, and political that can shed light to the inner worlds of this region.

5. Conclusion

Developing the premise of the invisible in cinema in greater detail, will act as the groundwork to analyze the cultural aspects of Balkan cinema. I'm not sure I will ever define the invisible, but this will give me a unique point of view in analyzing the works from the Auteur of the Balkan region. This point of view can lead to greater insight on the culture, the intrapersonal relationships, and the spirituality of the region. It will give us a glimpse of the Balkan “soul”.

References

- 1 Bresson R, Notes on the Cinematographer. Quartet Encounters 1986.

- 2 Fields S, Screenplay. Dell Publishing 1980.
- 3 Nizhny V, Lessons with Eisenstein. Da Capo Press 1979
- 4 Lumet S, Making Movies. Vintage Books 1995.
- 5 Director's Series, Tisch School of the Arts Graduate Film Program, New York University, 1991.
- 6 Horton A, The Films of Theo Angelopoulos: A Cinema of Contemplation. Princeton NJ: Princeton University Press 1997.
- 7 Corpus Media training of professional producers, European Media programme, June 2002.
- 8 Goodell G, Independent Feature Film Production. St. Martin's Griffin 1998.
- 9 Mulvey L, Visual Pleasure and Narrative Cinema. Screen 1975; 16 (3): 6-18.
- 10 43rd Annual Thessaloniki Film Festival, Balkan Survey, November 2001
- 11 Schrader P, Transcendental Style in Film: Ozu, Bresson, Dreyer. Da Capo Press 1972.
- 12 Barthes R, Camera Lucida. London: Fontana 1980.
- 13 Tarkovsky A, Sculpting in Time, Reflections on the Cinema. New York: Alfred A. Knopf 1987.
- 14 Florensky P, Iconostasis. St. Vladimir's Seminary Press 1996.
- 15 Iordanova D, Cinema of Flames: Balkan Film, Culture and the Media. London: BFI Publishing, 2001.

Modulation of IOR as a Function of SOA Value

Fenia Kalogeropoulou¹, Ana Vivas², Peter Woodruff³

¹SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, fkalogeropoulou@seerc.info

²City College, Dept. of Psychology, 24 Prox. Koromila, Str, 54622 Thessaloniki, Greece, vivas@city.academic.gr

³The University of Sheffield, Dept. of Clinical Psychiatry, Northern General Hospital, The Longley Centre, Norwood Grange Drive, S5 7JT, UK, p.w.woodruff@sheffield.ac.uk

Posner [1] developed an exogenous cue paradigm to investigate the biphasic effect of facilitation and inhibition in the orienting of attention. In this paradigm attention is first drawn to a peripheral location by the cue and then voluntarily removed to respond to the target. Reaction times (RTs) are faster to targets at the cued location (facilitation) than to the uncued when the time interval (SOA) between the onset of the cue and the target is short, and slower when the time interval is long (inhibition). This later effect is called inhibition of return (IOR) and reflects a bias of selective attention against returning to previously explored locations [2]. The aim of the present study was to investigate the change from facilitation to inhibition in a group of healthy adults. In addition, we were interested in the distribution of data around the cross-over point where early facilitation is replaced by inhibition. In order to do so we employed the single-cue paradigm [1] with five different time intervals between cue and target (SOA) (100ms, 140ms, 230ms, 350ms, 450ms) in twenty-eight healthy adults. We observed the expected validity and IOR effect, that is RTs were faster in cued locations in short SOA conditions and faster in uncued locations in long SOA conditions. Furthermore, 10% of the participants did not show the IOR effect, 15% and 25% changed from facilitation to inhibition at the 230ms SOA and at the 350ms SOA respectively, and the 50% of the participants displayed IOR in the long (450ms) SOA. This rather delayed appearance of IOR may be explained in terms of shifting strategies when using a short range of SOA [3]. Finally, we aim to apply the above procedure in patients with schizophrenia in order to compare patients' distribution of IOR with our results and investigate further the delayed onset of IOR effect in the disorder [4].

Keywords

Facilitation, IOR, schizophrenia.

1. Introduction: Modulation of IOR as a function of SOA value

Selective attention to locations and/or object in the space facilitates processing of relevant stimuli at the selected location or/and inhibits processing of irrelevant features [1] [2] developed a widely used experimental paradigm to investigate facilitatory and inhibitory processes of visuospatial selective attention. Typically, in a spatial cueing task an exogenous -non informative- cue (e.g., an abrupt onset) is employed to attract attention to one of two peripheral locations where a target is likely to appear. When the interval between the onset of the cue and the target (SOA- stimulus onset asynchrony) is short (around a 100ms), the expected facilitatory effect is found. That is, participants respond to the target faster at the valid location relative to the invalid location. However, studies have shown that when the SOA is longer than 250ms, participants detect the target faster in the invalid location as compared to the valid location. This phenomenon has been termed inhibition of return (IOR) [5], and it has been proposed to reflect a bias of selective attention against returning to the already -non-informative- explored location [2].

This biphasic effect of exogenous spatial cues, has been proposed to reflect attentional processes. Thus, facilitation first arises from the automatic capture of attention by the peripheral cue. On the other hand, IOR appears at longer SOAs, because attention is biased away from already explored locations in order for targets to be detected more efficiently at novel locations [6] [7]. However, other authors have proposed that both effects early facilitation and later IOR are independent and reflect the action of different mechanisms [8] [9]. Thus, some studies have reported inhibition without prior facilitation [8] then IOR reflects a change to the sensory level of the attentional system this is not very clear. In their study, [8] manipulated the temporal overlap between the cue and the target and found that IOR occurred only in long SOAs when there was no temporal overlap. On the other hand, facilitation was present only in short SOAs, where there was a temporal overlap between the cue and the target. Finally, the biphasic effect (facilitation followed by IOR) occurred only in the fixed cue duration condition, and when there was a temporal overlap in short but not in the long SOA condition. The researchers concluded that facilitation and IOR, following a spatial cue, are independent and are related to stimuli that signal attentional shifts such as the temporal properties of cue and targets. If the covert orienting of attention was biphasic in nature and if facilitation preceded inhibition, then IOR should arise as a consequence of the withdrawal of attention from the cued location. The current experiments showed that IOR can occur without prior facilitation so the two processes are independent [8].

Further evidence in support of this hypothesis is found in the study by Wright and Richard [10] which investigated the relation between cue and target-location predictability and IOR. The authors hypothesized that at short SOAs (100ms or less), the spatial cue would always lead to facilitation regardless of its predictability whereas at longer SOAs (300ms-400ms), IOR would not occur following informative cues. Indeed, the results of the study agreed with the authors' expectations: that is facilitation was present at short SOAs regardless of cue predictability: whereas IOR was presented only when the cue did not contain information about the location of the target. The above results suggest that facilitation and IOR are two separate processes [8] [9], with facilitation being purely reflexive and IOR depending on participants' beliefs about the location of target presentation [9] [10].

In relation to this debate, many studies have investigated what are the boundary conditions of IOR. That is, what are the conditions that modulate the onset of IOR [3]. For instance, it has been found that inhibitory effects with non-predictive spatial cues appear later when the task is more difficult [11] [12]. Thus, as in the seminal study of Posner & Cohen (1984), other studies have showed that facilitation usually is replaced by inhibition when the cue-target SOA is longer than 300 ms in a detection task. However Lupiañez and collaborators [11] [12] also showed that IOR appears later when participants are asked to perform a discrimination task. In a different study, [13] asked participants to detect either a perceptually degraded target or a high luminance target in a typical IOR paradigm. Similarly to Lupiañez and collaborators, he found that IOR was delayed by 25ms in the perceptually degraded target condition relative to the standard one. Thus, in general these studies [13] [11] suggest that the difficulty in detecting targets affects the allocation of attention, and support the idea that IOR does not simply reflect automatic, reflexive processes but involves high-level attentional process.

In addition to task difficulty, recent studies have shown that the onset of IOR is also modulated by the range of SOAs [3]. For example, [3] manipulated the range of the cue-target SOAs in a typical spatial cueing task. Specifically, they had three different range of SOAs, a short one which included SOAs values of 100, 250 and 400 ms, a middle range with included SOAs values of 100, 300 and 500 ms, and a long range which included SOAs values of 100, 400 and 700 ms. The authors found significant facilitation at the shortest SOA for all three SOA range. However, the onset of inhibition was modulated by the range of SOA, that is IOR emerged earlier with longer ranges of SOAs. [3] concluded that their results could be explained in terms of the participant's strategies to shift voluntary attention back to the center. That is, in the short SOA range participants may have delayed the shift of attention from the cued location in order to detect efficiently the target.

To conclude it is still not clear whether early facilitation and IOR are independent from each other or are the two ends of the biphasic effect. Furthermore, it seems that the onset of IOR is modulated by other experimental manipulations such as task difficulty and range of SOAs and is not purely a

reflexive automatic phenomenon. The aim of the present study was to further investigate the time course of facilitatory and inhibitory effects of non-informative exogenous cues in a group of healthy adults. Although, as we mentioned above many studies have already addressed this issue [3], none of them have looked at individual variance in the time at which facilitation is replaced by inhibition. A better understanding of how individual differ in the time course of inhibition, may be crucial for understanding the pattern of impaired IOR in patients with schizophrenia. Thus, we manipulated the cue-target SOA (100ms, 140ms, 230ms, 350ms, 450ms) in a typical single-cue IOR paradigm [2].

2. Method

Participants. Twenty-eight healthy adults participated in the study. They ranged in age from 18 to 32, with a mean of 24,7 years. 19 participants were recruited from the staff of the Psychiatric Clinic of 'Agia Fotini' in Larissa (Greece). The remaining 9 participants were undergraduate students of the University of Larissa. All of them participated voluntarily in the research. None of the participants had a history of mental illness, drugs or/and alcohol abuse.

Stimuli and Apparatus. Stimuli were presented on a colour monitor (VGA) of an IBM/PC compatible computer, and responses were recorded through the compute keyboard. The software used for creating and running the experiment was E-Prime. The stimuli were the two boxes and the target was a white square presented inside one of the boxes. Participants had to press the spacebar as soon as they saw the target.

Procedure. The experiment was explained verbally to the participants. Each trial began with a fixation point (a cross) that was presented in the middle of the computer's screen for a 1000 ms. Participants were instructed to look all time at the fixation point. Then two peripheral boxes appeared for 1000ms. One of the peripheral boxes became thicker for 50ms (the cue), and after a further time interval of 50ms, 90ms, 180ms, 300ms or 400ms, the target appear inside of one of the lateral boxes. The target was presented during 2000 ms or until a response was made (see Figure 1).

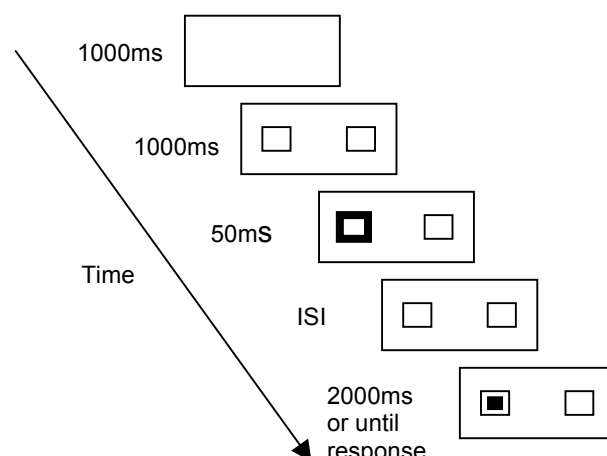


Figure 1 Sequence of events in a valid trial in experiment 1

Participants run one practice block of 16 trials and two experimental blocks of 300 trials each. In each experimental block, the target was present on 80% of the trials (260 trials), and it was absent on the remaining 20% of trails (40 trials). There were 52 trials for each SOA condition. On half of these trials (26 trials), the target appeared at the valid location, and at the invalid location on the other half.

Results. The mean of the median correct response times and standard deviation are shown in Table 1. Response times were submitted to a repeated measures analysis of variance (ANOVA) with SOA (100ms, 140ms, 230ms, 350ms, 450ms) and target location (valid and invalid) as the within subject factors. There was a marginally significant effect of target location $F(1, 27)=3.85$, $p = 0.06$, that is response times were faster for the cued condition (372ms) relative to the uncued (377ms) condition.

SOA	100	140	230	350	450
Cued	386 (73)	371(71)	356(66)	359(55)	388(64)
Uncued	399 (76)	386(77)	366(73)	367(81)	370(59)
	-13	-15	-10	-8	18

Table 1 The mean of the median correct RTs and standard deviation as a function of SOA and target location in Experiment 1

Also, there was a significant main effect of SOA $F(4, 108) = 8.89, p < 0.05$. Furthermore, there was a significant SOA by target location interaction $F(4, 108) = 8.89, p < 0.05$. The analysis of the interaction showed a significant effect of location for the 100 ms, the 140 ms and the 450 ms SOA, $F(1, 27)=10.33, p < 0.05$ RTs, $F(1, 27)=15.03, p < 0.05$ and $F(1, 27) = 18.42, p < 0.05$, respectively. That is, response time was faster at the valid locations (386ms and 371ms) relative to the invalid locations (399ms and 386ms respectively) for both the 100 and the 140 ms SOA, respectively. However, response times were faster at the invalid location (370ms) relative to valid locations (388ms) for the 450 ms SOA. Furthermore, 10% of the participants did not show IOR effect at all, whereas 15% and 25% of the participants showed inhibition at the 230ms and 350ms SOA respectively. The remaining 50% of the participants displayed IOR in the 450 ms SOA (see Appendix).

3. Discussion

As the results indicated 50% of the participants changed from facilitation to inhibition only to the long SOA (450ms) value, while 10% didn't show IOR at all. Only, 15% and 25% of the participants displayed IOR at a shorter SOA (230ms and 350ms respectively). Generally, these results are in line with previous findings [2] that suggest that facilitation is replaced by inhibition at about 300ms interval between the cue and the target. The rather delayed appearance of IOR at the 450ms instead at the 300ms as expected [2] [11] may be explained in terms of shifting strategies when the SOAs range is relatively short [3]. As it seems there are individual differences in the time course of IOR as not all participants maintained engagement in the cued location. The fact that 60% of the participants (50% showed IOR in 450ms SOA and 10% didn't show at all) the facilitatory effect of the cue persisted longer, indicated that exogenous and endogenous modes of allocation of attention are interactive and a percentage of participants endogenously maintained attention longer at the valid location. The above results suggest that IOR is not only reflexive (as if facilitation) but rather it depends on task demands and participants expectations [10]. Importantly, our study shows that there are individual differences in the deployment of attention when using non-informative exogenous cues. That is, not all participants showed inhibition at the same SOAs values. Future research may elucidate what are the specific factors that account for individual differences in the onset of IOR.

These results have also important implications for interpreting the impaired pattern of IOR observed in patients diagnosed with schizophrenia [4] [14]. Many studies have shown delayed IOR in these patients, however, the findings are inconsistent since other studies have reported intact IOR using a different procedure [15][16] [17]. Fuentes and Santiago employed a second central cue to withdraw attention from the periphery, and found the typical IOR effect in patients diagnosed with schizophrenia. The second central cue could have speeded up the development of IOR and overcome the delayed and blunted effect that had been reported previously [18].

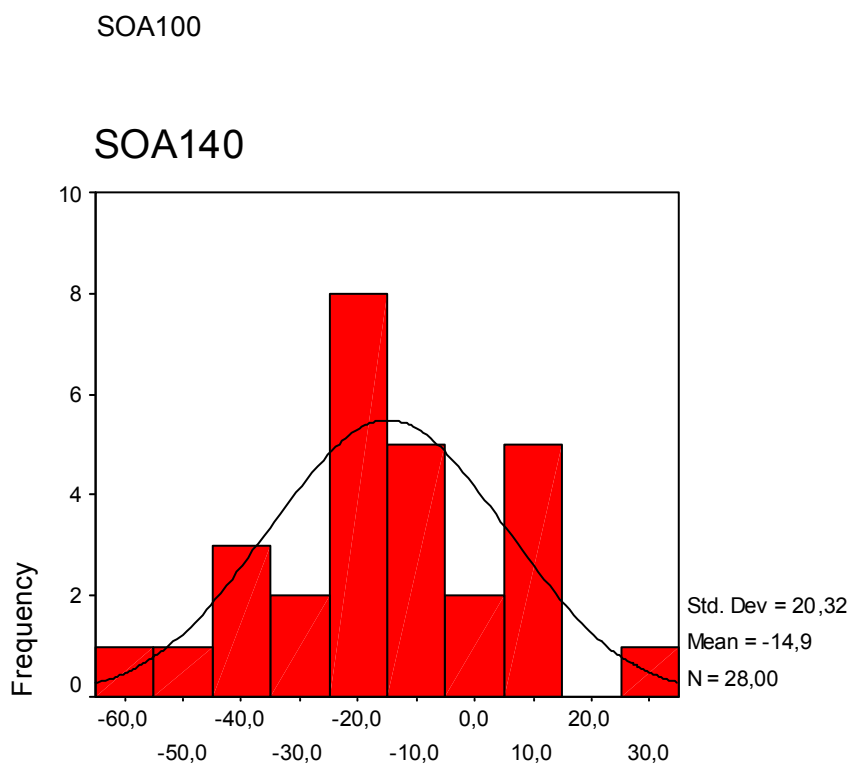
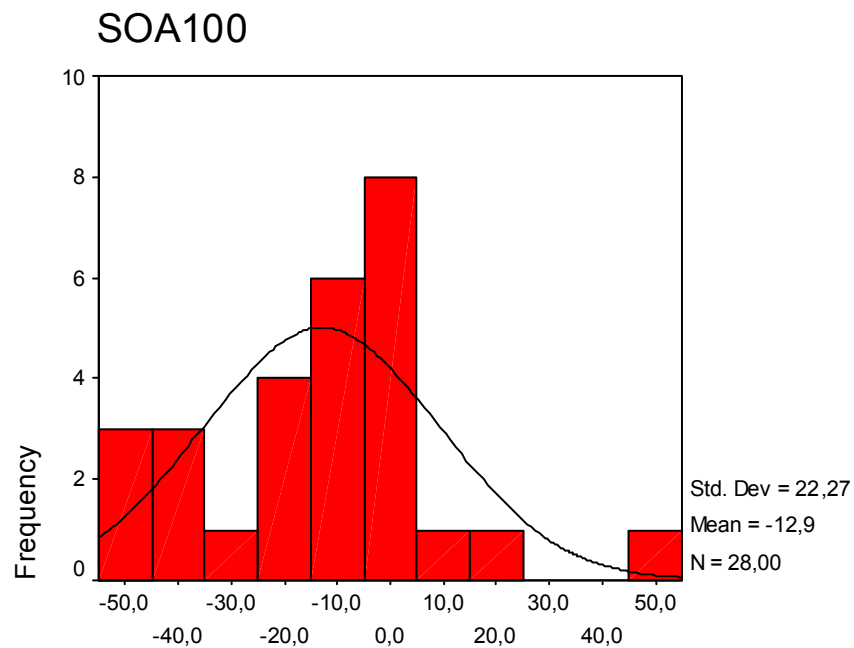
Thus, in future studies we aim at investigating the distribution of patients diagnosed with schizophrenia using a similar procedure to the one employed in the present study. It may be that in addition to the general impairment of delayed IOR onset, the patients may show a different distribution as compared to the healthy adults. Furthermore, these potential differences may be also modulated by other factors such as medication and symptomatology.

References

- 1 Posner M I. Orienting of Attention. *Quarterly Journal of Experimental Psychology* 1980; 32: 3-25
- 2 Posner M I & Cohen Y. Components of visual orienting. (in H. Bouma & D. Bouwhuis Eds.) *Attention and Performance* London: Erlbaum 1984
- 3 Cheal M & Chastain G. Timing of facilitatory and inhibitory effects of visual attention, *Visual Cognition* 2002; 9(8): 969-1002
- 4 Huey E D & Wexler B E . Abnormalities in rapid automatic aspects of attention in schizophrenia: Blunted Inhibition of Return. *Schizophrenia Research* 1994; 14: 57-63
- 5 Klein, R M. Inhibition of return. *Trends in Cognitive Science* 2000; 4: 138-147
- 6 Maylor E A & Hockey R. Effects of repetition on the facilitatory and inhibitory components of orienting in visual space. *Neuropsychologia* 1987; 25: 41-54
- 7 Posner M I & Petersen S E. The attention system of the human brain. *Annual Review of Neuroscience* 1990; 13: 25-42.
- 8 Maruff P, Yucel M, Danckert J, Stuart G & Currie J. Facilitation and inhibition arising from the exogenous orienting of covert attention depends on the temporal properties of cues and target. *Neuropsychologia* 1999; 37(6): 731-744
- 9 Tassinari G, Aglioti S, Chelazzi L, Peru A & Berlucchi G. Do peripheral non-informative cues induce early facilitation of target detection. *Vision Research* 1994; 34: 179-189
- 10 Wright R D & Richard C. Location cue validity affects inhibition of return of visual processing. *Vision Research* 2000; 40: 2351-2358
- 11 Lupianez J, Milan, EG, Tornay, FJ, Madrid, E & Tudela, P. Does IOR occur in discrimination tasks? Yes, it does but later. *Perception and Psychophysics* 1997; 59: 1241-1254
- 12 Lupianez J, Milliken, B, Solano, C, Weaver, B & Tipper, S P. On the strategic modulation of the time course of facilitation and inhibition of return. *Quarterly Journal of Psychology* 2001; 54A: 753-773
- 13 Castel A D, Pratt J, Chasteen A L. & Scialfa C T. Examining task difficulty and time course of IOR: detecting perceptually degraded targets. *Canadian Journal of Experimental Psychology* 2005; 59(2): 90-98
- 14 Gouzoulis-Mayfrank E, Heekeren K, Voss T, Moerth D, Thelen B & Meincke U. Blunted inhibition of return in schizophrenia. Evidence from a longitudinal study. *Progress in Neuro-Psychopharmacology & Biological Psychiatry* 2004; 28: 389-396
- 15 Fuentes L J, Boucart M, Alvarez R, Vivas A B & Zimmerman M A. Inhibitory processing in visuospatial attention in healthy adults and schizophrenic patients. *Schizophrenia Research* 1999; 40: 75-80
- 16 Fuentes L J & Santiago E. Spatial and semantic inhibitory processing in schizophrenia. *Neuropsychology* 1999; 14(1): 134-140
- 17 Sapir A, Henik A, Dobrusin M. & Hochman E Y. Attentional asymmetry in schizophrenia. Disengagement and inhibition of return deficits. *Neuropsychology* 2001; 15(3): 361-370
- 18 Larrison-Faucher A, Briand, KA & Sereno, AB. Delayed onset of inhibition of return in schizophrenia. *Progress in Neuro-Psychopharmacology & Biological Psychiatry* 2002; 26: 505-512.

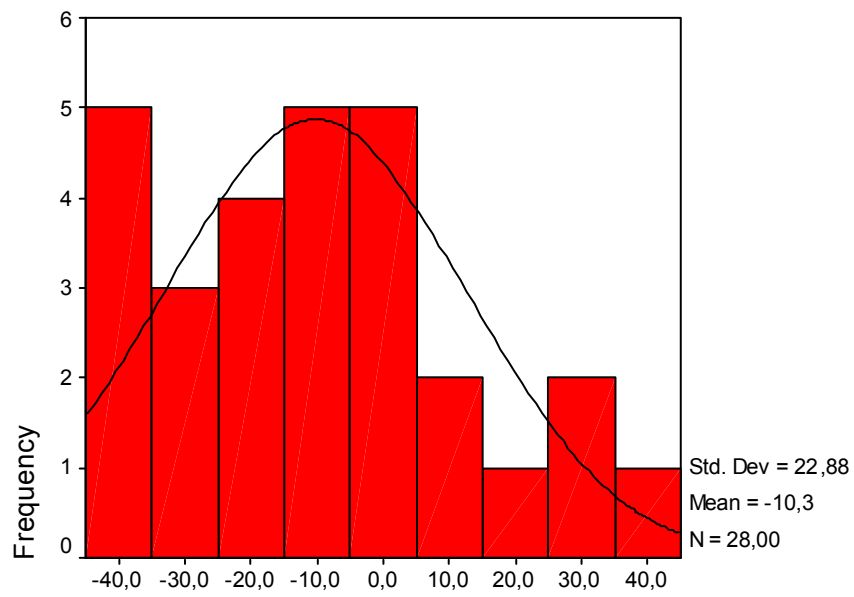
Appendix

Frequencies of participants' IOR effect according to the SOA value



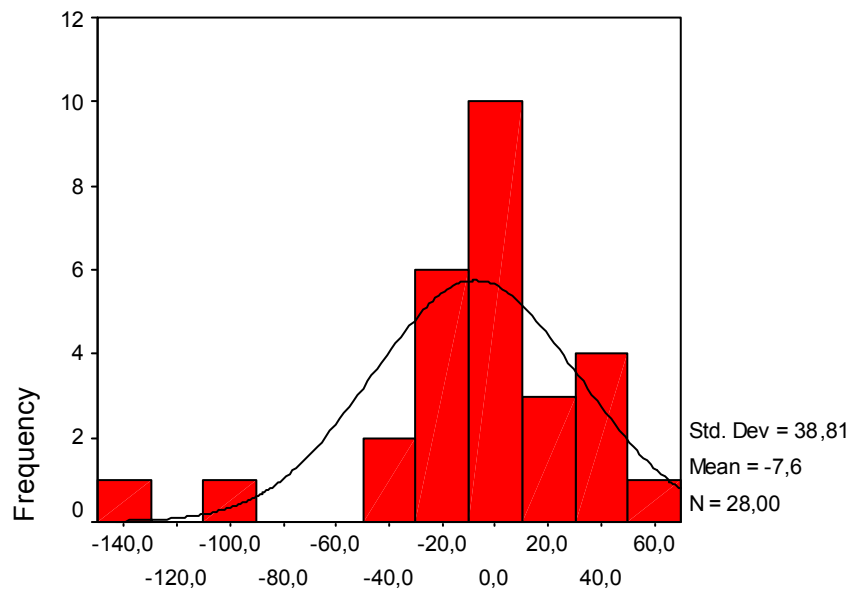
SOA140

SOA230

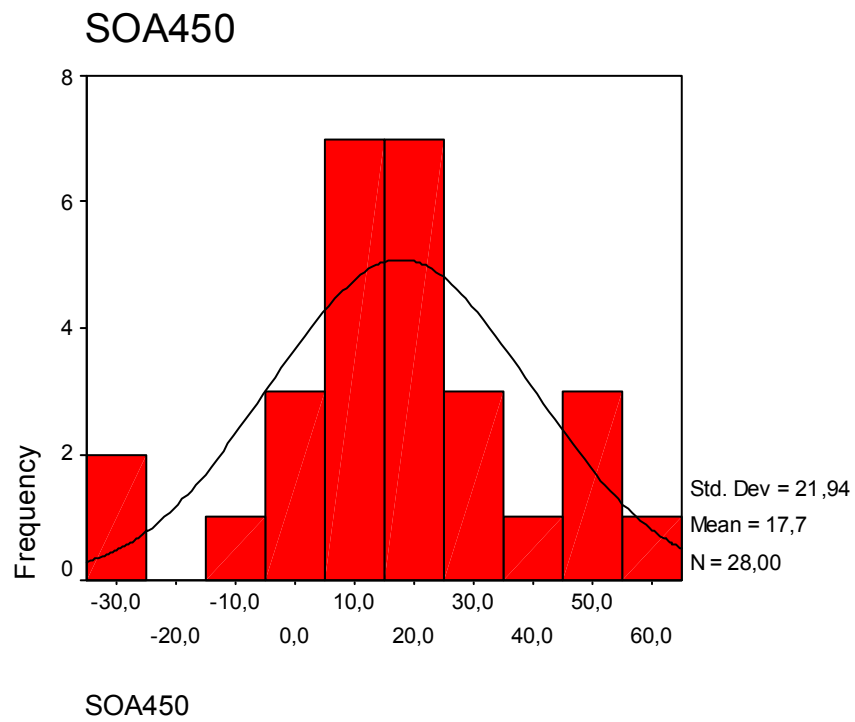


SOA230

SOA350



SOA350



A Preliminary Study of Greek Adolescents' Estimates of Smoking Prevalence: Beyond the Effects of False Consensus

Lambros Lazuras¹, Angelos Rodafinos², J. Richard Eiser³

¹SEERC, 17 Mitropoleos Str, 546 24 Thessaloniki, Greece, llazuras@seerc.info

²CITY College, Dept. of Psychology, 24 Prox. Koromila Str, 546 22 Thessaloniki, Greece, rodafinos@city.academic.gr

³The University of Sheffield, Dept. of Psychology, Western Bank, Sheffield, S10 2TN, UK, C.Eiser@sheffield.ac.uk

Subjective beliefs of smoking prevalence may play an important role in the development of adolescent smoking behaviour. Previous research has shown that smoking prevalence is often misperceived, mainly due to the effects of false consensus. The present study aimed to explore this assumption in a context where smoking prevalence is comparatively high, and prevention efforts against youth smoking are weak. A cross-sectional school-based survey was conducted with junior high-school students (N = 240, Mage = 13.7 years). Measures included perceived peer and adult smoking prevalence, demographic characteristics, smoking status, intentions to smoke, and frequency of observing other people smoking at school and leisure time public places. While smokers perceived greater smoking prevalence than non-smoker adolescents, the role of close friends who smoked and observing other students smoking at school were more important than individual smoking experience. Other significant predictors of prevalence estimates included senior grade in school, and female gender. The results also highlight the potential effects of availability heuristics on adolescents' frequency judgments of smoking. Findings are discussed in terms of the importance of normative processes in youth perceptions of smoking, and the re-examination of existing health education programs and anti-smoking regulations in schools and public places in Greece.

Keywords

Adolescent smoking, Greece, prevalence estimates, social norms.

1. Introduction

Despite recent decreases in tobacco use, smoking among youth still remains a major public health concern. Given the dose-dependent relationship between smoking and associated health problems, researchers strive to identify the predictors of smoking initiation and accordingly inform prevention strategies. In this quest, the importance of normative beliefs (i.e., how normative smoking appears to be in the adolescent's environment) has been highlighted. Such information is likely to provide cues to adaptive behaviour, and adolescents seem to follow them as an effect of social interaction [1-4]. Research has shown that adolescents who perceive high prevalence of smoking among peers and/or adults have more chances to become smokers [5-8]. Indeed, adolescents' subjective estimates of smoking prevalence among peers have been found to increase the risk for smoking [9-10]. While research appears supportive of a positive association between prevalence estimates and actual smoking in adolescence, relatively few studies have examined the psychosocial variables associated to those beliefs.

1.1 False consensus in adolescents' beliefs of smoking prevalence

Being a smoker and having friends who smoke are likely to influence adolescents' estimates of smoking prevalence. Research [11] has shown that frequency estimates of smoking behaviour can be shaped on the grounds of false consensus, as adolescents tended to draw inferences about the prevalence of peer smoking based on their own smoking experiences [see also 5]. More recently, though, Unger and Rohrbach [12] doubted the effects of false consensus, and argued that adolescents' prevalence estimates are largely determined by social influence in the form of best friend and classmates' smoking. The researchers further argued that the nature of false consensus effect could be more complex than assumed by Sussman and colleagues [11] in that adolescents may attend to their best friends' behaviour (and not necessarily their own) to draw inferences about smoking prevalence. Additionally, Unger and Rohrbach [12] found that female gender was a significant predictor of adolescents' judgments of smoking prevalence, but more recent studies [13] failed to replicate this association. It might be useful to re-examine whether adolescents' beliefs of smoking prevalence associate with gender differences.

1.2 The impact of distal variables on normative beliefs about smoking

At a more distal level, direct exposure to other people's smoke may as well affect the normative beliefs of adolescents. Research has repeatedly shown that teacher and peer smoking at school largely determine individual risk to initiate smoking [14-17]. Accordingly, prevalence estimates can be formed on the basis of observing other people smoke, since frequently observed behaviours tend also to be perceived as highly normative [1]. Unger and Rohrbach [12] argued that depictions of smoking as a normative behaviour (e.g., on TV or movies) can inflate adolescents' beliefs of smoking prevalence. Similarly, observing people smoke on school property may as well affect those estimates. This effect may extend beyond the school environment. Hence it may be worth examining whether frequent exposure to smoking in places where adolescents are likely to spend their leisure time affects perceived prevalence of tobacco use.

1.3 The role of cultural norms: The Greek case

Although researchers have elaborated on the impact of various psychosocial variables on estimates of smoking prevalence, the influence of cultural context has been undervalued. Cultural norms and pressures against (or for) smoking may perpetuate into people's beliefs about the prevalence of cigarette use. In this case, cognitive biases other than false consensus may influence adolescents' prevalence beliefs. Proponents of the availability heuristic assert that people's frequency judgments of events or behaviours depend on the availability of occurrences of the event or the behaviour being judged in memory [18-20]. It might seem reasonable to argue, therefore, that in a context where smoking appears rather prevalent and normative, instances of smoking are more frequent. Consequently, people's judgments of smoking prevalence may reflect reliance on availability heuristics, and not only the effects of false consensus beliefs.

Despite the increasing rates of smoking prevalence, Greece is probably among the countries with the weakest efforts to prevent smoking in EU, lacks national health surveys, and is considered the "paradise" for tobacco advertising [21- 23]. Also, there is scarce research in terms of identifying the correlates of tobacco use, and implementing evidence-based prevention interventions [24-25]. In addition, in a recent report, Labiris and colleagues [26] noted that school-based prevention programs in Greece appear rather ineffective in preventing smoking among adolescent students. At the same time, the international media describe Greece as the "last sanctuary for smokers", where anti-smoking rules are hardly complied with by the citizens and the respective authorities [27-29]. Greek adolescents, therefore, grow up in an environment where smoking is prevalent, prevention efforts appear rather weak, and the odds for biased estimates of smoking prevalence are increased. Evidence concerning the patterns by which subjective estimates of smoking prevalence differ from actual smoking rates, and whether these differences reflect some sort of cognitive bias can inform subsequent tobacco control policies.

1.4 The present study

The purpose of the present study was to explore the normative beliefs of Greek adolescents regarding perceived prevalence of smoking in peer groups and adults. Specifically, the study aimed to assess prevalence beliefs, and compare them with actual smoking rates. Smoking prevalence among 13-year-olds in Greece as derived from an international report of the Health Behaviour of School Children Survey [30] was used as a reference of peer-group smoking, together with self-reported current smoking of students in the participants' school. Actual rates of adult smoking derived from a recent publication of the World Health Organization [31]. A secondary aim of the study was to examine the psychosocial antecedents of prevalence beliefs.

2. Method

2.1 Participants

A total of 240 junior-high school students from a large northern Greek city participated in the study. This sample consisted of three student cohorts (1st, 2nd, and 3rd junior-high grades) aged between 12 and 17 years ($M = 13.6$, $SD = 1.1$; 45.9% males and 54.1% females). Participation was voluntary and students were informed about the purposes of the study and that they were free to withdraw at their will. Following the guidelines of ethical conduct in behavioural research, informed consent was obtained from the school principal, and parents/caregivers of the students.

2.2 Measures

The measures used in the present study were derived from a 33-item youth smoking survey that was developed and used in a larger project on the psychosocial correlates of youth smoking in Greece. Smoking status was measured on an 11-item nominal scale asking students to report their experiences with cigarettes. Items ranged from "No, I have never tried smoking" to "Yes, I currently smoke on average X cigarettes a day" where respondents had to fill in the number of cigarettes smoked. The question "how much do you want to be a smoker" was used to measure respondent's intentions to smoke in one month, one year, and five years' time respectively. Responses were assessed on a four-point scale ranging from "very unlikely" (= 1), to "very likely" (= 4).

Observation of other people smoking on school property was assessed with a three-item four-point continuous scale (from 1 = "never", to 4 = "very often"). A similar scale (from 1 = "almost never/never", to 4 = "almost always/always") was used to assess frequency of exposure to other people's smoking in typical leisure settings, such as cafes, bars/clubs, and internet stations. Open-ended questions were used to identify the number of close friends who smoked.

Similarly, open-ended questions asked participants to estimate the proportion of smokers in peer cohorts and among adults in Greece, and state the expected rate of smokers in 5 and 20 years' time. A typical item was "From 0-100%, how many people in your age group in Greece do you think are smokers?" and "From 0-100%, how many smokers do you believe will be in Greece in 5 years?"

2.3 Design/Procedure

A school-based, cross-sectional design was used to determine the associations between a set of psychosocial factors and adolescents' prevalence beliefs of peer group and adult smoking. Followed by the school principal, one of the researchers visited classes during regular teaching hours, gave instructions, and administered questionnaires. The class teacher was also present during this process, but together with the school principal agreed not to interfere with the procedure. Completion of the survey lasted approximately 15 minutes.

3. Results

3.1 Descriptive data

Mean scores and standard deviations of the study's variables are presented in Table 1. Most students (67.1%) were non-smokers, 27% reported having smoked at least one puff but not smoking now, and 5.9% reported current smoking. Reports of current smoking status appear representative of this age group, as the prevalence of current smoking among 13-year-olds in Greece is 6% (see Currie et al., 2004). Smoking status differed by grade in school, as senior students reported significantly higher rates of smoking than junior ones ($\chi^2 = 9.22$, $df = 2$, $p < .05$). No significant differences in smoking status were observed between male and female adolescents.

	M	SD	%
Gender			
Male	-	-	45.9
Female	-	-	54.1
Grade			
1st grade	-	-	32.9
2nd grade	-	-	36.3
3rd grade	-	-	30.8
Smoking status			
have smoked at least one puff, but not smoking now	-	-	27
Current smokers	-	-	5.9
never tried smoking	-	-	67.1
Intentions to smoke in			
1 month	1.10	0.42	-
1 year	1.12	0.41	-
5 years	1.27	0.61	-
Number of best friends who smoke	0.75	1.43	
Observation of smokers at school			
teachers	2.71	0.94	-
other students	2.34	1.00	-
Parents	1.32	0.64	-
Observing other people smoking in			
Cafes	3.64	0.91	-
restaurants	2.95	1.04	-
bars/clubs	3.23	1.34	-
Internet stations	2.74	1.28	-
Prevalence estimates of			
peer group smoking	43.93	23.42	-
adult smoking	75.00	15.32	-
Forecasts of future smoking prevalence			
5 years	76.71	17.31	-
20 years	80.10	22.10	-

Table 1 Study Variables

Note. Grades 1st through 3rd in the Greek junior high-school correspond to the ages of 12 – 14 years old.

The descriptive statistics that follow are indicative of the normative nature of smoking within the Greek context. More than half of the students (57.6%) reported that they frequently (often/very often) observed their teachers smoking on school property. Observing other students and parents smoking at school was less frequent (38.3% and 5.8% respectively). In leisure settings, most students reported frequent (often/almost always) exposure to other people's smoke in cafes (90.7%), bars/clubs (80.7%), and internet stations (59.4%).

3.2 Inferential statistics

3.2.1 Actual vs. perceived smoking prevalence

Using a series of one-sample t-tests, comparisons were made between perceived and actual smoking among peers and adults. Mean scores of perceived and actual prevalence are presented in Table 2. Significant differences were found between perceived and actual smoking prevalence for peer group ($t_{\text{HBSC}} = 24.9, p < .001$; $t_{\text{school prevalence}} = 25.1, p < .001$), and adult smoking ($t = 37.5, p < .001$).

	Perceived Prevalence %	Actual Prevalence %
Peer group smoking	43.9	6.0a 5.9b
Adult smoking	75.0	37.6c

Table 2 Mean Differences between Perceived and Actual Prevalence of Smoking

Note. Prevalence rates were derived from: (a) the Health Behaviour of School Children (HBSC) survey (Currie et al., 2004), (b) self-reports of current smoking of students in the participants' school, (c) the WHO (2003) report on smoking in EU countries.

3.2.2 Differences in perceived prevalence by smoking status, gender, and grade

A 2x2x3 MANOVA was used to determine the effects of smoking status, gender, and grade on perceived prevalence of smoking. Regarding the effects of smoking status, "non-smokers" were defined as those who never tried smoking, and "smokers" as those who had smoked at least one puff. Differences in mean prevalence estimates were significant only for peer group smoking. Significant main effects were observed for grade (Wilks' $\lambda = 0.91, F(4, 410) = 4.75, p = .001$) and smoking status (Wilks' $\lambda = 0.96, F(2, 205) = 3.67, p < .05$), but not for gender. Specifically, adolescent smokers reported higher prevalence estimates than non-smokers ($M_{\text{smokers}} = 51.7, M_{\text{non-smokers}} = 40.3$), and the same pattern was observed between senior and junior students ($M_{\text{1st graders}} = 36.5, M_{\text{2nd graders}} = 43.1, M_{\text{3rd graders}} = 53.2$). The interaction effects between the independent variables were not significant ($p > .05$).

3.2.3 Predicting prevalence estimates from personal and social factors

Personal factors included gender, grade, current smoking status, and intentions to smoke in the future. Social factors reflected best friends' smoking behaviour and observing other people smoke at school and leisure time public places. Using the enter method multiple regression models were generated to determine the independent variables that were significantly associated with prevalence estimates. The adjusted R^2 statistic was used to explain variance between predictor and criterion variables. A significant model emerged for prevalence estimates of peer group smoking ($R^2 = .29, F(14, 174) = 5.28, p < .001$), explaining 24.2% of the variance. Gender ($\beta = .272, p < .001$), grade in school ($\beta = .190, p = .007$), number of best friends who smoke ($\beta = .332, p < .001$), and frequency of observing other students' smoking at school ($\beta = .182, p = .019$) were the only significant predictors. Tolerance levels for these variables ranged from .50 and .86, thus reducing the odds for multicollinearity [32-33]. Prediction of prevalence beliefs regarding adult smoking was not significant ($p > .05$).

4. Discussion

Notably, adolescents' estimates of smoking prevalence among peers and adults were far beyond the actual rates of smoking, and the estimated rates are expected to escalate over the following five and 20 years. Although both personal and social factors significantly affected prevalence beliefs, it is important to note that the observed effects occurred in already inflated estimates. For instance, regardless of whether students or their best friends smoked or not, prevalence estimates were grossly overestimated, suggesting that smoking is perceived as highly normative. According to the assumptions of norm theory [34], adolescents seem to base their statistical estimates of smoking prevalence on what they perceive to be normal or unsurprising, or, in terms of heuristic biases, to information that is readily accessible from memory [18, 35]. This effect substantiates the argument that peoples' thoughts of statistical relationships are prone to error [36].

Botvin and colleagues [5] noted that adolescents who perceived that half or more than half of all adults or peers were smokers were at a greater risk for subsequent tobacco use. In a similar vein, Chassin and colleagues [6] argued that adolescents who overestimate the prevalence of smoking among peers and adults in general, might also feel strong pressures to smoke. It is reasonable, thus, to argue that unless those beliefs are changed, adolescents are at high risk for smoking initiation. In addition, those who really wish to stay smoke-free will probably perceive greater pressures against their intentions to abstain from cigarettes. The anticipated increase in future smoking prevalence that was reported in this study may further maximize perceived pressures to abstain, as smoking seems to be gradually portrayed as a standard behaviour. This latter effect may also influence beliefs of social acceptance, and desired societal norms around smoking. Previous research has shown that both of these variables are important antecedents of smoking behaviour [3, 8, 10, 14]. In this respect, the role of prevalence beliefs becomes more important than previously assumed, as there can be both direct and indirect effects on adolescents' decisions to initiate smoking.

A potential explanation of the present findings is that adolescents do not have sufficient knowledge concerning smoking prevalence in their country. To a certain extent, this can be attributed to the absence of national-based surveys of health behaviour [22], which would provide accurate and reliable information on health habits. An alternative explanation relates closely to research on the effects of availability heuristics on frequency judgments [see 18, 20, 37]. Specifically, within a cultural context where adolescents report frequent visibility of teacher smoking at school, and of other people smoking in leisure time settings, instances of smoking behaviour might seem more prominent than non smoking ones. In the face of frequency judgments of smoking, in turn, "smoking" instances are more easily retrieved from memory, and thus more likely to affect the judgment process. This effect can be exaggerated by dissemination of information that highlights the smoking behaviour of Greek people. Specifically, the fact that Greeks are labelled as "European Champions" in heavy smoking [see 24, 38] may create a consensus, and rather biased belief that the vast majority of Greeks *are* heavy smokers.

In either case, such misperceptions should normally be eliminated by school-based health education programs. Extending Labiris and colleagues' position [26], the present findings may reflect the inability of existing health education curricula in Greece to effectively tap aspects of tobacco use and prevalence, and highlight the need for interventions focusing on normative beliefs. More research is needed, though, to determine whether the observed effects are common among Greek youth, or merely a reflection of some sort of cognitive bias shared by adolescents in the specific high-school sample.

Regarding the effects of independent variables on smoking prevalence estimates, multivariate analyses of variance indicated that the social projection assumption is plausible, as smokers perceived higher prevalence than non-smokers. This finding is in accordance with previous research showing that adolescents base their estimates of smoking prevalence on their own smoking experiences [11].

Nonetheless, after controlling for the effects of covariates, regression analyses yielded a rather different finding. Specifically, female gender, grade, number of best friends who smoked, and frequent exposure to other students' smoking at school, were the only significant predictors of estimates of peer group smoking. This latter finding is in accordance with Unger and Rohrbach [12] who also reported a significant association between these variables and perceived prevalence of peer group smoking.

Therefore, the false consensus assumption seems plausible when adolescents are asked to estimate the prevalence of peer smoking, yet from a social norms perspective. That is, adolescents indeed tend to overestimate the smoking behaviour in peer cohorts, and these estimates appear to be based on the smoking behaviour of their close friends and other students at school, rather than their own experiences with cigarettes.

4.1 Limitations and strengths of the study

Although the findings appear interesting in terms of exploring adolescents' perceptions of smoking prevalence, there are some limitations that need to be mentioned. Firstly, it was a preliminary study, which used a convenience sample (a single school), and this limits the generalizability of the findings to other school populations. In addition, the regression models explained only a relatively modest amount of variance (>30%). Perhaps additional measures, such as parental or sibling smoking at home might have led to a better prediction of prevalence estimates. Future studies may overcome these problems by using more representative samples and more extensive measures of descriptive norms.

Notwithstanding these limitations, the present study also had some strong points. Firstly, this report is the first ever done in Greece concerning adolescents' normative beliefs about smoking. Secondly, the findings suggest that within a normative context for smoking, cognitive shortcuts like availability heuristics may play an important role in shaping adolescents' judgments of the frequency of smoking behaviour. This is important as it provides the potential for an integrative approach concerned with the interplay between cultural context, judgment biases, and the formation of normative beliefs around smoking. Finally, the results provide guidelines for school-based prevention initiatives and tobacco control policies in Greece.

4.2 Policy-relevant outcomes

In terms of school-based prevention there seems to be a need for programs aiming to challenge adolescents' current beliefs of smoking prevalence, with a special focus on females and students who are about to transfer to senior high-school. According to previous studies [see 5] if overestimation beliefs remain unchanged, the odds for smoking initiation are likely to increase as tobacco use is represented as a highly normative behaviour that most peers and adults engage in. Moreover, these beliefs may adversely influence the motivation of adolescents who are keen to abstain from cigarettes. It is important to protect non-smokers by creating an environment where smoking abstinence is portrayed as normative rather than deviant or minority behaviour. This way, normative pressures against becoming a smoker are minimized, and adolescents willing to stay smoke-free will be more motivated to utilize their intentions [3].

In terms of tobacco control policy-making, there also seems to be a need to enforce strict rules concerning smoking bans in schools. Even though the present report showed that prevalence estimates remained unaffected, there is evidence that teacher smoking significantly contributes to the development of smoking habits among students [17]. Moreover, the finding that observing other students smoking at school influenced significantly prevalence estimates of peer group smoking strengthens the argument towards strong smoking restrictions within the school setting [39-40]. Besides, there is empirical evidence supporting the association between visibility of smoking in school premises and individual risk for smoking initiation [16].

Similarly, a strict anti-smoking policy has to be enforced in public places where adolescents choose to spend their leisure time. It is striking to find out that the vast majority of adolescents in this study reported frequent exposure to other peoples' smoking in leisure settings that should already have separate sections for smokers, or any other policy aiming to protect non-smokers and young people from being exposed to environmental tobacco smoke [38, 40-42]. From a theoretical standpoint, strict anti-smoking policies are likely to convey anti-smoking norms more effectively, and thus maximize the joint impact of both subjective (i.e., what should be happening), and descriptive (i.e., what is happening) norms on behaviour [1-2].

4.3 Conclusion

Overall, the present findings suggest that future research on adolescent smoking should explore the link between cultural norms and judgments of smoking frequency. In addition, smoking prevention initiatives in Greece should consider the effects of normative processes on youth perceptions of smoking, and their relations to subsequent uptake. Recalling Leventhal and Cleary's [43] argument, it is doubtful whether smoking prevention interventions will be effective at all, unless social norms against smoking are changed. Perhaps this assertion is still timely in a country where norms around smoking appear far weaker than in most EU countries and the US.

References

- 1 Cialdini R B. Crafting normative messages to protect the environment. *Cur Dir Psychol Science* 2003; 12: 105-9.
- 2 Cialdini R B, Reno R R, Kallgren C A. A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *J Pers Soc Psychol* 1990; 58: 1015-26.
- 3 Petraitis J, Flay B R, Miller T Q. Reviewing theories of adolescent substance use: organizing pieces in the puzzle. *Psychol Bull* 1995; 117: 67-86.
- 4 Rimal R N, & Real K. Understanding the influence of perceived norms on behaviours. *Com Theory* 2003; 13: 184-203.
- 5 Botvin G J, Botvin E M, Baker E, Dusenbury L, Goldberg C J. The false consensus effect: predicting adolescents' tobacco use from normative expectations. *Psychol Rep* 1992; 70: 171-8.
- 6 Chassin L, Presson C C, Sherman S, Corty E, Olshavsky R. Predicting the onset of cigarette smoking in adolescents: a longitudinal study. *J App Psychol* 1984; 14: 224-43.
- 7 Conrad K M, Flay B R, Hill D. Why children start smoking: predictors of onset. *Brit J Addiction* 1992; 87: 1711-24.
- 8 Tyas S L, Pederson L L. Psychosocial factors related to adolescent smoking: a critical review of the literature. *Tob Control* 1998; 7: 409-20.
- 9 Ellickson P L, Bird C E, Orlando M, Klein D J, McCaffrey D F. Social context and adolescent health behaviour: does school-level smoking prevalence affect students' subsequent smoking behaviour? *J Health Soc Behav* 2003; 44: 525-35.
- 10 Wiium N, Torsheim T, Wold B. Normative processes and adolescents' smoking behaviour in Norway: a multilevel analysis. *Soc Sci Med* 2006; 62: 1810-8.
- 11 Sussman S, Flay B R, Dent C W, Mestel-Rauch J, Johnson C A, Hansen W B. Adolescent non smokers, triers, and regular smokers' estimates of cigarette smoking prevalence: when do overestimations occur and by whom? *J App Soc Psychol* 1998; 18: 537-51.
- 12 Unger J B, Rohrbach L A. Why do adolescents overestimate their peers' smoking prevalence? Correlates of prevalence estimates among California 8th-grade students. *J Youth Adol* 2002; 31: 147-53.
- 13 Lai M K, Ho S Y, Lam T H. Perceived peer smoking prevalence and its association with smoking behaviours and intentions in Hong Kong Chinese adolescents. *Addiction* 2004; 99: 1195-1205.
- 14 Eisenberg M E, Forster J L. Adolescent smoking behaviour: measures of social norms. *Am J Prev Med* 2003; 25: 122-8.
- 15 Flay B R, Hu F B, Siddiqui O, Day L E, Hedeker D, Petraitis J, Richardson J, Sussman S. Differential influence of parental smoking and friends' smoking on adolescent initiation and escalation of smoking. *J Health Soc Behav* 1994; 35: 248-65.
- 16 Leatherdale S T, Brown K S, Cameron R, McDonald P W. Social modelling in the school environment, student characteristics, and smoking susceptibility: a multi-level analysis. *J Adol Health* 2005; 37: 330-6.
- 17 Poulsen LH, Osler M, Roberts C, Due P, Damsgaard MT, Holstein BE. Exposure to teachers smoking and adolescent smoking behaviour: analysis of cross sectional data from Denmark. *Tob Control* 2002; 11: 246-51.
- 18 Kahneman D, Slovic P, Tversky A. *Judgement under uncertainty: heuristics and biases*. Cambridge University Press 1982.

- 19 Tversky A, Kahneman D. Availability: a heuristic for judging frequency and probability. *Cog Psychol* 1973; 5: 207-32.
- 20 Williams K W, Durso F T. Judging category frequency: Automaticity or availability? *J Exp Psychol* 1986; 12: 387-96.
- 21 Costa J, Mossialos E. Which policies effectively enable smoking cessation? Evidence from the European Union. *Int J Soc Econ* 2006; 33: 77-100.
- 22 Hupkens C L H, van den Berg J, van der Zee J. National health interview surveys in Europe: an overview. *Health Policy* 1999; 47: 145-68.
- 23 Kokkevi A, Terzidou M, Politikou K, Stefanis C. Substance use among high school students in Greece: outburst of illicit drug use in a society under change. *Drug Alcohol Depend* 2000; 58: 181-8.
- 24 Koumi I, Tsiantis J. Smoking trends in adolescence: report on a Greek school-based, peer-led intervention aimed at prevention. *Health Promot Int* 2001; 16: 65-72.
- 25 Lazuras L, Rodafinos A. Survey of smoking attitudes in grade school children in Greece: a preliminary study. *Eur Addict Res* 2006; 12: 20-4.
- 26 Labiris G, Voutsinas A, Niakas D. Preliminary evaluation of the school-smoking-prevention policy in Greece. *Eur J Public Health* 2005; 15: 329-30.
- 27 Agence France Presse. Smokers in Greece stand their ground. <http://www.tobacco.org/newsfeed/country/greece.rss> January 2006
- 28 Burlington Free Press. Greetings from Greece. <http://www.tobacco.org/newsfeed/country/greece.rss> December 2005
- 29 Vardavas C I, Kafatos A. Greece's tobacco policy: another myth? *Lancet* 2006; 367: 1485-6.
- 30 Currie C, Roberts C, Morgan A, Smith R, Settertobulte W, Samdal O, Rasmussen V B. Young people's health in context - Health Behaviour in School-aged Children (HSBC): International report from the 2001/2002 Survey. WHO Regional Office for Europe 2004.
- 31 World Health Organization. WHO European country profiles on tobacco control. WHO Regional Office for Europe 2003.
- 32 Brace N, Kemp R, Snelgar R. (2000). *SPSS for psychologists: a guide to data analysis using SPSS for Windows*. Palgrave 2000.
- 33 Lewis-Beck M S. *Applied regression: an introduction*. Sage 1980.
- 34 Kahneman D, Miller D T. Norm theory: comparing reality to its alternatives. *Psychol Review* 1986; 93: 136-53.
- 35 Sherman S J, Judd C M, Park B. Social cognition. *Ann Rev Psychol* 1989; 40: 281-326.
- 36 Eiser J R. *Social judgment*. Open University Press 1990.
- 37 MacLeod C, Campbell L. Memory accessibility and probability judgments: an experimental evaluation of the availability heuristic. *J Pers Soc Psychol* 1992; 63: 890-902.
- 38 Centres for Disease Control and Prevention. Tobacco information and prevention sources-Europe-Greece. <http://www.cdc.gov/tobacco/who/greece.htm> January 2006.
- 39 Moore L, Roberts C, Tudor-Smith C. School smoking policies and smoking prevalence among adolescents: multilevel analysis of cross-sectional data from Wales. *Tob Control* 2001; 10: 117-23.
- 40 Wakefield M A, Chaloupka F J, Kaufman N J, Orleans C T, Barker D C, Ruel E E. Effect of restrictions on smoking at home, at school, and in public places on teenage smoking: cross sectional study. *BMJ* 2000; 321: 333-37.
- 41 The Council of European Union. Council's recommendations on the prevention of smoking and on initiatives to improve tobacco control. *J Eur Union* 2003; 22: 31-4.
- 42 Wold B, Currie C, Roberts C, Aaroe L E. National legislation on school smoking restrictions in eight European countries. *Health Promot Int* 2004; 19: 482-8.
- 43 Leventhal H, Cleary P D. The smoking problem: a review of the research and theory in behavioural risk modification. *Psychol Bull* 1980; 88: 370-405.

