

ΤΩΝ  
ΚΟΙΝΩΝΙΚΩΝ  
ΕΠΙΣΤΗΜΩΝ

ΤΟ ΒΗΜΑ

- *C. Ménard* : The new institutional approach to institutions
- *J. Backhaus-G. Schaefer* : Striving to be member of the club
- *N. Rodger* : The military revolution at sea
- *N. Kyriazis-M. Zouboulakis* : The Economics of Sea Power
- *K. Aravossis-E. Koutsiana* : Program evaluation methodologies
- *M. Tietzel-Ch. Müller* : The peculiar economics of doping in sports
- *G. Halkos* : Testing the existence of an Environmental Kuznets Curve

# SOCIAL SCIENCE TRIBUNE

Quarterly Scientific Review

Published since 1989

*Special Issue in English*

## ECONOMY AND SOCIETY

Volume10

Issue 37

UNIVERSITY OF THESSALY  
GREECE

# SOCIAL SCIENCE TRIBUNE

## Contens

<b><i>Claude Ménard:</i></b> The new institutional approach to institutions: concepts, methods, results .....	5
<b><i>Jürgen G. Backhaus, Günther G. Schaefer:</i></b> Striving to be member of the club: when institutions become variables and outcomes become given .....	27
<b><i>Nicolas Rodger:</i></b> The military revolution at sea .....	59
<b><i>Nicholas Kyriazis, Michel Zouboulakis:</i></b> The Economics of Sea Power. Property rights arrangements and institutional change in Elizabethan England .....	77
<b><i>Konstantinos Aravossis, Efrosini Koutsiana:</i></b> Program evaluation methodologies – A comparative assessment .....	97
<b><i>Manfred Tietzel, Christian Müller:</i></b> The peculiar economics of doping in sports .....	117
<b><i>George Halkos:</i></b> Testing the existence of an Environmental Kuznets Curve for sulfur using panel data models .....	135



## The new institutional approach to institutions: concepts, methods, results

Claude Ménard\*

### *Abstract*

*New Institutional Economics has developed considerably over the last twenty years, qualifying as a «progressive research program». This program is based on a hard core of concepts that apply to well identified objectives: The alternative modes for organizing transactions, the institutional environment in which they are embedded, and the interaction between these dimensions. Transaction costs play a major role in this approach. This paper examines why it is so, it deals with several controversial issues, and it provides examples of recent breakthroughs in economics, but also in history and in the political sciences.*

---

\* 1. Professor of Economics, Université de Paris I (Pantheon-Sorbonne), ATOM, ménard@univ-paris1.fr. This paper is a slightly modified version of a paper published in French in *Cahiers d'Economie Politique* nr. 44, 2003. I would like to thank the participants of the seminar «What have we learned about institutions?» as well as seminars given at the Ecole Normale Supérieure (Paris), the Université de Sceaux, the Erasmus Universiteit (Rotterdam) and the University of Thessaloniki for their comments on previous versions. I would most particularly like to thank Bertrand Bellon, Christian Bessy, Joëlle Farchy, Maarten Janssen, Uskali Maki, Philippe Mongin, Isabelle This-Saint Jean, Bernard Walliser and Michel Zouboulakis for their comments. I retain, of course, the unique responsibility for any errors and opinions expressed in this paper.

## 1. Introduction

As its name describes, the new institutional approach is built on and revolves around the subject of institutions. The main proposal developed in the following pages can be summarized as follows: the new institutional program is a «progressive research program» as defined in Lakatos (1976), containing a hard core of concepts relating to specific, well-identified objects and whose field of application goes deeper as compared to the initial questions, while rapidly expanding to new questions (Canguilhem, 1968, Introduction).

The corollary proposal, one which I will of course try to justify, consists of supporting that under the apparently somewhat overgrown view of the research program, a powerful unifying element transpires: namely, the approach in terms of transaction costs.

Why should the accent be immediately put on the transactional approach? For one simple but fundamental reason, noted very early on by Ronald Coase (1937), who has since continued to evoke this essential point (Coase, 2001): transactions are at the core of the economic activity. In fact, without efficient mechanisms to ensure transactions, we cannot take advantage of the benefits of the division of labor and the resulting organization of the production would lose its entire meaning. By transaction, I mean, following Williamson (1985, p. 1), *the transfer of «rights to use» goods or services between separable technological units*. The idea of «rights to use» that I introduce here, thus modifying the definition of Williamson, seems to me to be most important. For example, the case of the transfer of property rights is in fact only one aspect, essential of course, of understanding market mechanisms. But the integrated company also organizes transactions internally; transactions that Demsetz (1991) prefers to call the transfer of «administrative rights». Furthermore, at a more general level, social methods exist for the transfer of rights other than property rights; for example, when a community systematically organizes a collective transfer of rights. Therefore, the definition that I have just proposed deserves general recognition, as it became more specific in accordance with the subjects studied. It combines the abstraction and precision expected of a key concept of a theory.

Further, the very idea of mechanisms that are essential for the organization of these transfers of rights also has a major consequence. If alternative mechanisms exist, then these mechanisms will have

relative costs that can be compared. Choosing the «right» mechanism thus becomes an essential part of the strategy of microeconomic units as well as a determinant factor of innovation and growth. Therefore, in this paper, I will put the accent on costs adopting almost exclusively an economist's approach to transactions and the related institutions. I insist on this point because of the importance of the distribution and influence of new institutional theories in numerous other social sciences, most particularly political sciences, history of economics and management theory<sup>1</sup>. While I sometimes make reference to some of these developments, the quick overview that I propose here is essentially related to researches in economics.

The article is organized into three major sections. Section two reexamines the main concepts used in the definition of the research topics of the neo-institutional program and the resulting consequences. Section three discusses the problems related to the methods that led a certain number of «orthodox» as well as «heterodox» economists to attack the neo-institutional program. Finally, the fourth section summarizes certain results that I consider particularly important, and raises questions that I think may nurture the development of the new institutional approach in the coming years. I will conclude drawing a certain number of lessons learned from that ongoing research program.

## 2. Key concepts of new institutional economics

It can be, of course, quite overwhelming to begin the discussion of the contributions of a theory with definitions. But this is almost inevitable since there must be a positioning of the new institutional approach with regard to alternative paradigms, which are partly complementary and partly competitive. Part of what differentiates the various institutional approaches is related to the subject the approach is devoted to.

### 2.1. The starting point

Initially, we may adopt a simple and classical distinction between «institutional environment» and «institutional arrangements», one proposed by Lance Davis and Douglas North (1971, pp. 6-7) and developed more precisely by North in 1990. Environment refers

back to the rules of the game, the political, social and legal rules that define and support the transactional activities of the actors, while arrangement refers back to the actor's method for using these rules, or more exactly, to the methods of organization of transactions within the scope of these rules. I believe that Coase, at his Nobel Prize conference, revealed a much more precise and satisfactory content of these arrangements, as he spoke of the «institutional structure of production» (Coase, 1991).

Therefore, we must note that the distinction proposed here progressively revealed two branches to the new institutional research program. A first branch, rather general, explores the nature and the role of institutions by highlighting their historical dimension (North, 1981; Greif, 1998; Aoki, 2001). Regarding this aspect, the concepts of *enforceability* (*ex-ante*) and *enforcement* (*ex-post*) play a key role: under an economic angle, an institutional environment is not characterized only by the production of rules and/or standards (hence the problem of the types of rules), but also and above all by the production of mechanisms destined to implement these rules and by their effective implementation. The analysis of these mechanisms is a key point of the new institutional program. Measuring their relative costs is also a major challenge.

A second branch is of a rather micro-economic nature. It is probably the most familiar one to economists, and relates to the study of the methods of organization of transactions, arbitration between these methods, and their compared efficiency, with a strong analytical dimension<sup>2</sup>. The concept of *incomplete contracts* plays an essential part here, for at least two distinct reasons.

Firstly, the idea that contracts play a structuring role in the organization of transactions allows for a precise translation of the relational dimension essential to the very concept of transaction. Secondly, the idea that practically all contracts are incomplete, naturally leads to the requirement of an in-depth analysis of the coordination mechanisms complementary to the contractual agreements (even implicit) as for example the form assumed by «command» (i.e., the role of hierarchy in firms), in the creation and allocation of resources.

## 2.2. Returning to the concept of institution

Let us return to the concept of institution that implies directing the analysis in the areas we have just indicated. I am aware that



there is no universally accepted definition of concept of institution. In a very restrictive and minimalist sense, an institution can be viewed as an agreement between two agents. For many, the institution is a balance between the strategies of the agents participating in a «game» (as defined within game theory, of course). Aoki recently pushed the analysis much further in this regard (Aoki, 2001). Others have even put the accent on the fact that an institution is concerned with the very rules that govern the game, in short, a sort of meta-agreement. In following the path indicated by North (1990), I would like to propose a more precise definition, whose main interest is to define the field of analysis and, through this, to better clarify the new institutional research program. By institution, I therefore mean *a set of rules that are durable, stable, abstract and impersonal, appearing in the laws, traditions and customs and embedded in the mechanisms that implant and implement, by consent and/or constraint, the methods of organization of transactions.*

This definition involves a certain number of points that I would like to emphasize.

The first point is that stability and durability are essential to the identification and characterization of the rules of the game that mark an institution. New Institutional Economics offer the hypothesis that there is a very weak variability in institutions over time, in particular as compared to the modes of governance (or organizational modes). An important consequence is therefore deduced, there is a strong differentiation of the time horizons of the branches of the new institutional program described above. Williamson proposed to think of institutions in secular terms while organizational arrangements are essentially intra-secular (Williamson, 2001).

Second point, these rules of the game are abstract and impersonal in the sense that they transcend not only the individuals but also the organizations. They transcend in that they are perceived as being non arbitrary, imposed in a non discretionary manner on well defined classes of agents or entities (Hurwicz, 1987). A rule that varies in line with the individuals to whom it applies cannot take its support from stable mechanisms and is quickly faced with problems of implementation. «Customized» rules and constraints are not interpreted as institutions by the agents to whom they apply, but according to the case, as privileges arbitrarily attributed to certain agents or as coercive or unjust measures. One of the key elements of the dynamic of institutions undoubtedly results from this tension.

Finally, institutions possess a normative character. They exist on-

ly through the mechanisms that define «the limited set of accepted alternatives in a society at a given moment» (North, 1986). From this viewpoint, institutions possess a double standard, of which results an obvious paradox. On the one hand, they definitely and significantly restrict the field of action of the individual agents and, due to this, require *enforcement* mechanisms. On the other hand, they permit development of the transactional activity. Analysis of the mechanisms destined to make effective the rules required by this transactional activity constitutes an essential component of the new institutional analysis and explains the role that authors such as Coase may have held in the development of the programs of «Law and Economics».

I furthermore proposed a characterization of these specific «micro-institutional» mechanisms (Menard and Shirley, 2002, chap. 1).

The analysis of these micro-institutions, for example the cases of «regulation» with regard to the general laws that govern them and to the constraints they imposed on the players of the game, seems to me to be a significant field of research. But, on the other hand, institutions may enlarge the fields of action of the agents by safeguarding their transactions. Many of North's studies (summarized in his book of 1981), but also in Alston and Libecap (1996) regarding property rights, illustrate this point. The events leading to the transition towards a centrally planned economic market constitute an *in vivo*, and distressing, experience of this key role of institutions.

Naturally, a very interesting question is formulated from the definition of my proposition above, and this question tends to take a non negligible place in recent literature: does it make sense to speak of private institutions? The new institutional response is «yes» and examples have been given and analyzed (see the series of articles of North et al. on the «Law Merchant», for example, Milgrom, North and Weingast 1989).

### 2.3. Some consequences of the new institutional design of institutions

An examination of the design of institutions to which I just referred, and that touches on just a part of the problems put forth, led me to clarify certain aspects that I consider crucial for future research.

Firstly, what role does the concept of transaction play in the proposed definition? This is rather central, since, as economists opera-

ting in the context of this design, we are mainly interested in institutions and the accompanying mechanisms («micro-institutions») that will facilitate transactions and reduce their costs. Hence, the key point of these mechanisms is *enforceability*, or the ability to enforce rules and procedures ex-ante for feasible implementation, and *enforcement*, or the necessary mechanisms to make these rules operational ex-post, in such a manner so as to secure the transactions of the agents. For example, the adoption of legal rules for the protection of trademarks must provide for the possibility of court action, but would not have any relevancy without the establishment of adequate administrative and legal mechanisms guaranteeing their effectiveness (for illustrations, see Menard, 1996; Menard and Valceschini, 1999).

Further, may we speak of efficient institutions? I know that I will create many disagreements on this point but I will defend a «YES» answer, in the sense that there is *comparative efficiency*. Certain institutions are more favorable than others for the development of a significant volume of transactions, hence the possibility of taking better advantage of the division of labor; and for the reduction of transaction costs, hence the possibility of higher accumulation. We may think of, for example, the rules and mechanisms governing the transfer of property rights in real estate, and the difference between countries regarding these rules and the associated costs. The hypothesis that I make, and that can be supported by comparative studies, therefore is that institutions exist that carry easy transactional abilities and, through this, permit taking better advantage of the division of labor. Alston proposes characterizing these institutions as «growth enhancing», which could be a more adequate characterization than the term «efficient». In any case, the central idea, that North (1971, 1981) put forth early on and that Greif (1993,1998) recently developed with very precise examples, remains the same: institutional mechanisms exist that, in the long run, have greater stimulating effects than others.

Finally, as a last remark, this does not mean that the efficiency criterion of institutions is the only one to be taken into account. On the one hand, we find here the old cleavage between «efficiency» and «equity». The reduction of transaction costs and the increase of volume, and eventually even quality of the transactions they permit, may very well be accompanied by a devastating amplification of the inequalities, or effects having catastrophic results for the environment. A certain number of political scientists and historians have

started to use the transactions concept to analyze all the exchange that takes place between efficiency and equity on a political level or to explain major historical bifurcations<sup>3</sup>.

On the other hand, the efficiency criterion says nothing about the stability of institutions: institutions that are perfectly inefficient in that they restrict the development of transactions or lead to a serious increase of their costs, may in fact remain in place, for a very long time (Greif, 1998).

#### 2.4. What about institutional arrangements?

Careful readers may have noted that in all of the preceding, I did not speak much on the other part of the new institutional program, the more micro-analytical part, that concerns what Davis and North (1971) called «institutional arrangements» and that gave rise to all the recent literature on the modes of governance (Williamson, 1996). I will refrain from developing this aspect, so as to focus the attention, in the assigned space, on the crucial question asked, relating to institutions in a general sense. We must however keep in mind that, in the approach developed here, the modes of organization definitely have institutional roots<sup>4</sup>. I will briefly come back to certain aspects of this question in the last section («Results»), if for no other reason than the necessity of clarifying the interaction between the institutional environment and these organizational modes, a point where the new institutional approach has also significantly progressed.

### 3. Method

One of the major difficulties that the economic analysis comes up against when it tackles institutions, relates to problems of methods. The question is in fact to know if the economist's classic tool box is adequate for this subject. The answer is complex because the technical tools and the underlying hypothesis fit together inextricably.

It is out of the question for me to elaborate extensively here on the problems of methods brought about by the new institutional program (and to which they come up against)<sup>5</sup>. But I believe it is important to say a few words because a large part of the critics of

«heterodox» as well as «orthodox» economists<sup>6</sup> with regard to this research program are developed from this angle.

### 3.1. A «standard» method...

In a sense, New Institutional Economics present no methodological originality concerning the research method. They submit to a classical logic revealed by the history of sciences, including social sciences. In fact, they are based on (1) a theory, that is to say a set of well defined questions and concepts built to analyze these questions; (2) models developed from these concepts, mobilizing the analytical instruments that allow to generate testable propositions on classes of well defined phenomena; and (3) tests, effectively destined to confront these propositions with data, either by the use of measure (including econometric tests) or by a recourse to simulation, either abstract (game theory) or applied (experimentation).

However, let me emphasize the first difference that is also a source of major difficulty, as much in the relationships with the standard approach as within the research program itself. The modelization intent mentioned above is accompanied in fact by a «plausibility» requirement concerning the underlying hypotheses. This leads the new institutional authors to systematically retain two key hypotheses: the hypothesis of bounded rationality and the hypothesis of opportunistic behavior. The first one comes up against the «neo-classical» sensitivity as it leads to a questioning of the postulate of maximization, which would render the use of the related tools less productive, if not impossible. The second greatly hinders the «heterodox» authors who view this as a continuation of the idea that the agents are essentially calculative, which would make them «a-sociable» and therefore abstractions that are not very pertinent for the analysis of the real social phenomenon. I do not pretend to have a satisfactory answer to these two «hindrances». But I would like to simply notice that these difficulties have not stopped the research program from progressing rapidly, as I will try to show in the next section. And I believe that we can be inspired by the history of sciences to nourish a certain confidence: as a general rule, when a pertinent question is put forth, it ends up by creating adequate analytical tools.

### 3.2. ...that also mobilizes original elements

We can find traces for this requirement for new methods for ex-

ploring with the use of non conventional techniques that complex issues rose by institutions and how they change over time and shape trade-offs among modes of organization, in at least two directions. Hence, in my opinion, taking this into account is essential if we hope to understand the progress made in the analyses of institutions and if we want to go further in regard to these topics.

The first direction is that of comparative analysis. We cannot proceed with the analysis of institutions in a purely axiomatic manner. The very identification of the rules of the game and the study of their application, of their impact on the performances of the organizational modes, and more generally, on the dynamic of organizations, implies marking the research terrain with points for comparison. Comparison does not mean description. The comparative studies undertaken by the new institutionalists are guided by a theory, the transactional approach. In the present state of the research, many works<sup>7</sup> use a mix of qualitative analyses and theoretical models that are basically «local», destined to account for the introduction and the application of certain rules.

A typical case seems to be that of a model developed by Milgrom-North-Weingast (1989) that accounts for the nature and the role of the «Law Merchant» in the trade fairs of the Middle Ages.

Also, Avner Greif is developing a whole series of models destined to account for the «enforcement» mechanisms and for their place in the development of Mediterranean commerce of the 11th, 12th, and 13th centuries. In another perspective, let me quote the works of McCubbins, Weingast, etc. regarding the impact of the political rules of a federated system as compared to a centralized system in the implementation of a transactional order.

A second direction concerns the use of case studies, a very controversial subject for numerous economists of the «mainstream». Case studies play an important role in updating and analysing the general rules of the game, but also in the study of institutional arrangements and modes of organization (an emblematic example is that of the buyout of Fisher Body by General Motors. See the controversial contributions to the *Journal of Law and Economics*, April 2001). Most of the authors working in the new institutional field do not seem to be suffering from the allergy that for obscure reasons, is quite widespread among economists, in regard to «case» studies serving as a basis for theoretical developments. The reason could be due to a misunderstanding. If we refer to a case study as descriptive research, then the new institutionalists share the scepticism of con-

ventional economists. But can we characterize as a purely descriptive monograph the study of the pins factory that has provided Adam Smith with a basis for the establishment of the concept of division of labor? Or the study of a baseball team, that led to the creation of the tournament theory? Or the analysis of the franchising of cable television in Oakland, that led to a profoundly new interpretation of arbitration between modes of organization? Would that be the case, we should reject a very substantial part of the theoretical developments in economics over the last two centuries.

## 4. Results

After having brought forth the conceptual background and the methodological problems, what additional knowledge, can the new institutional program claim to have contributed to the analysis and the comprehension of the nature and the role of institutions? Here also, it is not appropriate to go ahead with an exhaustive inventory, considering the abundance of accumulated results. If we look back on the last twenty years of this research program, the summary would seem quite impressive.

Standard economics make no mistake about this, preempting more and more questions and results. I will be content with recalling three areas of the program that seem to me to be most fruitful so far.

### 4.1. Analysis of organizational modes (or arrangements) and arbitration between these arrangements

A first set of contributions, undoubtedly the one that has progressed the most rapidly and most particularly since the 1980's, concerns «institutional arrangements», or, according to the terminology that I have rather favored until now, organizational modes. It is the program, well known and renowned, impelled by Williamson (1975). This domain of new institutional researches contains precise and testable predictions, therefore refutable. I would like to emphasize three points in particular.

The first is to recall the impressive number of studies and tests on integration and quasi-integration, studies based on the concept of *specific assets*, and, although to a much lesser extent, on the *uncer-*

*tainty* variable, surrounding transactions. As numerous authors have said (see, for example, Joskow, 1991; 2002), the analysis of integration is a «success story» of the new institutional program. The employed concepts allow us to better understand how firms' boundaries are determined, and how the arbitration is carried out between alternative institutional arrangements. Some comparative tests have also shown that the new institutional explanation is «performing» particularly well as compared to alternative explanations (Poppo and Zender, 1998; Rindfleisch and Heide, 1997; Whinston, 2001).

A second contribution that I would like to emphasize concerns the very distinct development, over the past ten years, of studies on an alternative class of arrangements, the *hybrid organizational forms* (sometimes characterized as «non standard» institutional arrangements). By this we refer to modes of governance based on agreements between legally autonomous entities, but sharing a subset of economic decisions, arrangements designed to preserve the incentive advantages of the market while implementing mechanisms to reduce opportunistic behavior. Examples, now considered classic, are networks among firms, franchises, producer's groups. While there are many earlier related works, the article that truly launched these studies was that of Williamson (1991). This author developed the idea, until then basically intuitive, of stable arrangements presented between the polar cases of the markets and the «hierarchies», therefore making considerably more complex the picture he had drawn in his book in 1975. A summary of the characteristics of these arrangements was proposed in Ménard (1997), and since then the studies have multiplied trying to learn more and better understand the contracts underlying the agreements, and more generally, coordination and incentive mechanisms that ensure sustainability (see, for example, Brousseau and Glachant (ed), 2002; Lafontaine and Reynaud, 2001; Ménard 2003).

On the other hand, there is a certain paradox on the micro-analytical side of the new institutional program: This lies in is the relative underdevelopment of the studies regarding the internal characteristics of integrated organizations. In fact, taking into account the importance attributed in 1937 to the «nature of the firm» by the founder of the new institutional movement, Ronald Coase, and taking into account that the first research works of Williamson<sup>8</sup>, related precisely, to the internal properties of integrated firms and the manner in which they determine their advantages as well as their limits with regard to the market, we could have expected a more rapid de-



velopment of researches on these issues. Hence, the progress in this area is still poor, as emphasized by numerous authors (see for example, the numerous contributions to the *Journal of Economic Perspective*, April 1998).

Therefore, we possess very few studies today relating to the internal mechanisms of firms that allow them to organize transactions internally in an efficient manner (Gibbons, 2001), or in the manner in which human asset specificity plays out on the organization of work and, therefore, the relative efficiency of a firm with regard to other institutional arrangements. Part of the explanation lies in the difficulty in obtaining data. But it is crystal clear that the analytical framework must also be developed. This is where the new institutional concerns join with those of economists closer to the central core of the mainstream (for example, Holmstrom, 1999).

#### 4.2. Analysis of a certain number of institutional mechanisms and their effects on the organization of transactions

A second area of fruitful research concerns the analysis of institutions themselves. The works of Douglass North have evidently played an important role in directing this dimension of the new institutional program. Here again, too many contributions should be taken into account. Let me focus on some of them.

Firstly, progress has been made concerning the comparative analysis of institutional mechanisms, which underlines their role in the differentiation of the paths of development because of their impact on the volume and the methods of organization of transactions. The political-legal dimension holds a particularly important place here, and the concept of «enforceability» plays a key role in these analyses that partly inspired the annual report of the World Bank (2001). To be more precise, I will give here three examples that are particularly significant: (1) the work of Avner Greif, comparing the political mechanisms established in Genoa and in Venice in the 12th and 13th centuries and leading to organizational choices and development paths that were strongly differentiated; (2) the research of Sokoloff, Engerman and Haber (2001) on the compared growth of the USA and Latin America; or (3) the works of Lee Alston et al. (1996) regarding the role of land property rights in the dynamics (and some of its failures) of the American economy.

We should also take note of the researches that have allowed for the identification and the analysis of the «efficient» (or inefficient) institutional rules, in the sense that they permit an increase in the volume of transactions and a reduction of their cost. I must particularly mention here the remarkable development of studies on the intersecting area of «Law and Economics» (in particular, contract rights and property rights), mainly initiated and piloted by Coase, within his editorial functions for the *Journal of Law and Economics*<sup>9</sup>.

Less known, particularly in Europe where these developments have not greatly advanced until now, but very dynamic is the new institutional movement developing at the intersection of economics and political sciences, for example regarding the nature and the role of a «federal» system, the relative advantages of the degree of centralization and the impact on the «enforceability» of the rules of the game. If an important part of these studies has been in regard to the American system (Weingast, Ferejohn, McCubbins), they are also being extended to other areas, for example, Brazil (Mueller) or Argentina (Spiller and Tommasini).

I would finally like to make note of a certain number of more formalized works, regarding mainly the question of institutional stability, for example, in the analysis of the factors that determine institutional equilibrium and the factors that lead from one equilibrium to another. The pioneering work of Hurwicz (1987), on «institutional design», that remained at a very general and abstract perception of institutions as a system for transmitting messages, gave way to much more precise researches on the conditions of stability of well defined rules of the game, whether they are related to norms (Aoki, in Ménard 2000; Aoki 2001), to political rules (McCubbins, 1999), or to social mechanisms relevant to the micro-economic organization of transactions (Ensminger, 1992).

#### 4.3. Analysis of interactions between institutional environment and modes of organization

The preceding naturally lead to highlighting other results, even more recent and, therefore, currently undergoing elaboration and deeper exploration, relating to the manner in which the institutional environment interacts with the organizational arrangements that are displayed within this context. The numerous new institutional studies of the last decade regarding the phenomenon of regulation/de-

regulation have allowed for advancement in this regard, in particular in three areas.

First of all, we now have better knowledge of the impact of the choice of rules of the game on the manner in which the modes of organization of transactions are determined and therefore the exchange between these modes as well as on the resulting performances. We may refer, for example, to the works of Joskow and Schmalensee (1997) for the manner in which State-owned enterprises in Russia have been delineated in the privatization movement, to the works of Joskow regarding the deregulation of the electrical sector (Joskow 1991b, 1998), to the studies of Levy and Spiller (1994) for the institutional transformation of the telecommunications sector or the studies coordinated by Shirley (2002) in the water sector. An interesting development concerns the use of experimental economics to test the behaviors related to the rules of the game (see for example, Staropoli, 2001).

A second series of interesting results on this subject relates to the analysis of the impact of the legal frameworks on the choice of the modes of governance and their performances. First of all, there is a series of studies relating to the manner in which property rights are defined as well as to the consequences resulting from the way the transactions are organized in such contexts. Good examples are provided in the works of Alston, Libecap and Mueller (1997) regarding American or Brazilian agriculture, or the works of Libecap and Wiggins (1985) regarding operations in the oil fields. Then, there are researches regarding the laws governing the operations of firms, for example, the laws concerning «corporations» and their consequences on the mode of governance (illustrated, for example, by Marc Roe, 2002). Finally, there are the analyses concerning the role of the legal system in the definition, elaboration and execution of contracts (I can refer here to the works of Alan Schwartz (1992), among others). All these researches converge, despite their diversity, in what they all emphasize, the articulation between legal rules and organizational choices based on these rules.

I would like to end this far from exhaustive review, by mentioning a series of works in the process of publication that put more and more emphasis on the central role of «micro-institutions», interspersed, on the one hand, between the global rules of the game marking the institutional environment, and on the other hand, the agents, the organizations or the contractual agreements that bind them. Converging studies (Levy and Spiller, 1994; Shirley (ed.),

2002; Menard and Shirley, 2002) show, in fact, the key role of these mechanisms in the organizational choices and in the performances observed once the choices are made. The objective of these researches is also to progressively identify the key *micro-institutions*, to show how they are closely linked to certain sectors of activity. A theory for intermediary institutions then takes shape, setting out the general rules of the game for the effective modes of organization of transactions in specific activities. It is these intermediary *institutions* that, in all probability, make all the difference regarding the efficiency of institutions comprising the global institutional environment and largely determine the differences in performance, explaining why the same mode of governance (sometimes the same firm) succeeds in one environment and fails in another.

## 5. Conclusion

I have presented here but a minute part<sup>10</sup> of the recent contributions of new institutional analysis to the knowledge of what institutions are and how they function and interact with the arrangements for which they serve as a framework and support. As I have shown, there is quite an abundance of contributions. Readers will conclude, I hope, that the new institutional program is effectively «progressive», that it continues to open up new fields and therefore deserves more discussion.

If I must retain a few unquestionably central points of these contributions, I would definitely put forth the concepts of transaction and transaction costs with all the analytical mechanisms that surround them. But I would also emphasize the analysis that is developed around the problems of «enforceability» and «enforcement», on the nature and the characteristics of incomplete contracts, and the now central idea in the new institutional program that there exists a significant variety of alternative organizational modes, among which a never ending arbitration takes place, an arbitration where both the rules of institutional environment are at play as well as the internal characteristics specific to these modes of organization. Finally, I would like to note the study of the conditions characterizing the efficiency and the stability of the institutional mechanisms.

Major problems obviously remain, of which I should have made a point. I will mention three of these points that I particularly feel

strongly about. (1) We have, as of yet, made little progress on the question of innovation, in particular regarding the nature and the articulation of organizational and institutional innovations. (2) We do not, or barely, understand the mechanisms for the implementation of incentives in their organizational environment and in their institutional dimension. (3) We have not advanced very far in the area of analyses of the agents' behaviors allowing us to go above the *ad hoc* hypotheses related to bounded rationality and opportunistic behaviors. In particular, we are still debating the famous hypothesis of rationality. Here, I believe we share a difficulty common to all economic theories, attributing a central position to the analysis of institutions. A major problem results: For lack of a more generalized theory, we have a tendency to produce localized theories, beyond the transaction concept that cannot include everything. Fortunately, we may draw an optimistic conclusion: there is still a lot of time left for more research into institutions.

## Notes

1. For a survey of the diverse contributions of the new institutional approach to these fields and some others, see Ménard, C. and M. Shirley (2003; forthcoming).
2. For overviews on the subject, see Williamson, 1985; Joskow, 1988; Crocker and Masten, 1996; Klein and Shelanski, 1995, Rindfleisch and Heide, 1997, Ménard 2003.
3. Let me mention, for example, the researches of Ferejohn, McCubbins, Weingast in political sciences, of Engerman, Nye, Sokoloff in history.
4. For developments on this point, see Ménard (1995).
5. For a more detailed approach, see Ménard, 2001. One can also find very interesting remarks, although from a totally different viewpoint, in Lafont 1999.
6. And other social scientists in that respect.
7. References are provided in the next section.
8. See, for example, his remarkable article of 1967 on «Hierarchical Control and Optimum Firm Size».
9. A concise presentation of these contributions can be found in the preface of Yves-Marie Morissette, a law scholar from Montreal (Introduction to Coase, 2000).
10. For a more complete view, see Ménard (2003) and Ménard and Shirley (2003).

## References

- Alston, L.J., Libecap, G.D., Mueller, B. (1997), «Violence and the Development of Property Rights to Land in the Brazilian Amazon». In Drobak, John and John Nye, *The Frontiers of the New Institutional Economics*, Academic Press, Chap. 7.
- Alston, L.J., Libecap, G.D., Schneider, R. (1996), 'The Determinants and Impact of Property Rights; Land Titles on the Brazilian Frontier', *Journal of Law, Economics and Organization*, 12 (1), 25-61.
- Aoki, M. (2000), «Institutional Evolution as Punctuated Equilibrium», in Ménard (ed), 2000., pp. 11-35.
- Aoki, M. (2001), *Towards a Comparative Institutional Analysis*, Cambridge (USA), MIT Press.
- Brousseau E., Glachant, J.M. (2002) *The Economics of Contracts*. Cambridge (UK): Cambridge University Press.
- Canguilhem, G. (1968), *Etudes d'histoire et de philosophie des sciences*, Paris, Librairie Vrin.
- Coase, R.H. (1991), «The Institutional Structure of Production». Alfred Nobel Memorial Prize Lecture in Economic Sciences. Nobel Foundation: 12 pages. Reproduit dans *American Economic Review*, 1992, 82(4): 713-19.
- Coase, R.H. (1999), «The Task of the Society». Opening Address to the Annual Conference of ISNIE, *ISNIE-Newsletter*, 2(2):Fall.
- Coase, R. (2000), *Le coût du droit*, Paris: Presses Universitaires de France. Introduction by Yves-Marie Morrisette
- Crocker, Keith J. et Scott E. Masten (1996) «Regulation and Administered Contracts Revisited: Lessons from Transaction-Cost Economics for Public Utility Regulation». *Journal of Regulatory Economics*, vol. 9.
- Davis, L.E., North, D.C. (1971) *Institutional Change and American Economic Growth*. Cambridge (UK): Cambridge University Press.
- Demsetz, H. (1991), «The Theory of the Firm Revisited», dans Williamson, Oliver et Sydney Winter (eds), *The Nature of the Firm*, New York :Oxford University Press, chap. 10, pp. 159-78.
- Engerman, S., Haber, S., Sokoloff, K. (2001), «Inequality, Institutions and Differential Paths of Growth among New World Economies», in Ménard C. (ed), *Institutions, Contracts and Organizations*, Edward Elgar Pub., pp. 108-35.
- Ensminger, J. (1992), *Making a Market*, Cambridge (UK), Cambridge University Press.
- Gibbons, R. (2001), «Why Organizations are Such a Mess (and What an Economist Might Do About It)», Working paper, MIT.
- Greif, A. (1993), «Contract Enforceability and Economic Institutions In Early Trade: The Maghribi Traders» *American Economic Review*, 83 (June): 525-48.

- Greif, A. (1998), «Historical and Comparative Institutional Analysis» *American Economic Review*, 88 (May), no. 2: 80-4.
- Holmstrom, B. (1999), «The Firm as a Subeconomy». *Journal of Law, Economics and Organization*, 15:74-102.
- Hurwicz, L. (1987), «Inventing New Institutions: the Design Perspective». *American Journal of Agricultural Economics*. 69 (2): 395-402.
- Joskow, P. (1991a) «The Role of Transaction Cost Economics in Antitrust and Public Utility Regulatory Policies». *Journal of Law, Economics and Organization*. 7 (Sp): 253-83.
- Joskow, P. (1991b), «Asset Specificity and the Structure of Vertical Relationships: Empirical Evidence», in Williamson, O., Winter, S. (eds.), *The Nature of the Firm*, Oxford, Oxford University Press, Chap. 8, pp. 117-37.
- Joskow, P. (1997), «Restructuring, Competition and Regulatory Reform in the US Electricity Sector». *Journal of Economic Perspectives*, 11(3): 119-38.
- Joskow, P. (2002), *Journal of Law, Economics and Organization*, 18.
- Joskow, P., Schmalensee, R. (1997), «Privatization in Russia: What Should be a Firm?», in Ménard C. (ed.), *Transaction Cost Economics: Recent Developments*, Edward Elgar Pub., chap. 4, pp. 86-126.
- Klein, P., Shelanski, H. (1995), «Empirical Research in Transaction Cost Economics: A Survey and Assessment», *Journal of Law, Economics and Organization*, 11(2):335-61.
- Lakatos, I. (1976), *Proofs and Refutations*. Cambridge (UK), Cambridge University Press.
- Laffont, J.-J. (1999), «Political Economy, Information and Incentives», *European Economic Review*, 43: 649-69.
- Lafontaine, F., Raynaud, E. (2000), «Creance residuelle et flux de rentes comme mecanismes incitatifs dans les contrats de franchise: complements ou substituts ?», *Revue d'economie industrielle*, no 92: 255-76.
- Levy, B., Spiller, P. (1994), «The Institutional Foundations of Regulatory Commitment», *Journal of Law, Economics and Organization* 10 (Fall): 201-46.
- Libecap, G., Wiggins, S. (1985), «The Influence of Private Contractual Failure on Regulation: The Case of Oil Field Unitization», *Journal of Political Economy*, 93: 690-714.
- McCubbins, M. (1999), Ménard, Cl. (1995), «Markets as Institutions vs Organizations as Markets: Disentangling Some Fundamental Concepts», *Journal of Economic Behavior and Organizations*, 28 (3).
- Ménard, Cl. (1996), «On Clusters, Hybrids and other Strange Forms. The Case of the French Poultry Industry», *Journal of Institutional and Theoretical Economics*, 152 (1).
- Ménard, Cl. (2001), «Methodological Issues in New Institutional Economics», *Journal of Economic Methodology*, 8(1):85-92.

- Ménard, Cl. (2003a) «The Economics of Hybrid Organizations». *Journal of Institutional and Theoretical Economics* (forthcoming)
- Ménard, Cl. ed. (2003b), *The International Library of the New Institutional Economics*, 7 Volumes, Edward Elgar Pub. (Forthcoming).
- Ménard, Cl., Shirley, M. (2002), «Reforming Public Utilities, lessons from Urban Water Supply in Six Developing Countries», The World Bank.
- Ménard, Cl., Shirley, M. (2003), *Handbook of New Institutional Economics*, Boston, Kluwer Academic Press (forthcoming).
- Ménard, Cl., Valceschini, E. (1999), «The Creation and Enforcement of Collective Trademarks», *Voprosi Ekonomiki*, Moscow, March: 74-87.
- Milgrom, P., North, D.C., Weingast, B. (1989), «The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs», *Economics and Politics*, 2:1-23.
- North, D.C. (1981), *Structure and Change in Economic History*, New York, Norton and Co.
- North, D.C. (1986), «The New Institutional Economics», *Journal of Institutional and Theoretical Economics*, 142(1):230-37.
- North, D.C. (1990), *Institutions, Institutional Change, and Economic Performance*, Cambridge, Cambridge University Press.
- North, D.C. (1991), «Institutions». *Journal of Economic Perspectives*, 5 (Winter): 97-112.
- Poppo, L., Zender, T. (1998), «Testing Alternative Theories of the Firm: Transaction Cost, Knowledge Based and Measurement Explanations for Make-or-buy Decisions in Information Services», *Strategic Management Journal*, 19: 853-77.
- Rindfleisch, A., Heinde, J. (1997), «Transaction Cost Analysis: Past, Present, and Future Applications», *Journal of Marketing*, 61 (October): 30-54.
- Roe, M. (2002), *Political Determinants of Corporate Governance: Political Conflict, Corporate Impact*. New York, Oxford: Oxford University Press.
- Schwartz, A. (1992), «Relational Contracts in the Courts: An Analysis of Incomplete Agreements and Judicial Strategies», *Journal of Legal Studies*, 21: 271-318.
- Shirley, M. (ed) (2002), *Thirsting for Efficiency*, Amsterdam, Elsevier Pub.
- Staropoli, C. (2001), *Conception de marches efficaces pour les secteurs dereglementes: le cas des marches de gros d'electricite*, *These de Doctorat*, Universite de Paris (Pantheon- Sorbonne).
- Whinston, M. (2001), «On the Transaction Determinants of Vertical Integration»
- Williamson, O. (1967), «Hierarchical Control and Optimum Firm Size», *Journal of Political Economy*. 75: 123-38.



- Williamson, O. (1975), *Markets and Hierarchies: Analysis and Antitrust Implications*, New York, The Free Press.
- Williamson, O.(1985), *The Economic Institutions of Capitalism*. New York, The Free Press.
- Williamson, O. (1996), *The Mechanisms of Governance*. New York, Oxford University Press.
- Williamson, O. (2001), «The New Institutional Economics: Taking Stock/looking ahead», *Journal of Economic Literature*, 37(3):595-613.





## Striving to be member of the club: when institutions become variables and outcomes become given

*Jürgen G. Backhaus\**, *Günther G. Schaefer\*\**

### *Abstract*

*Successful, less successful and unsuccessful examples of system transformation processes leading to or trying to lead to stable democratic market economies are plentiful.*

*These developments pose a challenge to economics and political science to develop theoretical concepts that can explain the dynamics of these system transformation processes and that can suggest strategies for a successful management transition and ensure stability. We propose to take the first steps towards developing such a conceptual framework by interpreting elements of relevant economic and political theory. We shall do this by first taking a critical look at relevant theoretical concepts in political science and economics before, in part III outlining in a rather tentative and general way one proposal for an integrated approach.*

We are currently witnessing a series of most remarkable political system transformations in many parts of the world. Developments in Eastern Europe and in the former Soviet Union are probably the

---

\* Prof. Dr. Lic. jur. Krupp Foundation Chair in Public Finance and Fiscal Sociology, Faculty of Economics, Law, and the Social Sciences, University of Erfurt, Germany.

\*\* European Institute of Public Administration, Maastricht.

most dramatic. The disintegration of the Soviet Union and the decline of its influence on Eastern European nations where new leadership, committed to the establishment of market economies and democratic systems of government, have assumed responsibility, dramatically illustrate the wave of fundamental change that is presently sweeping Europe and which has its –probably less successful counterparts– in other parts of the world, particularly in Latin-America and South-Africa. These developments are fundamentally changing the world order that has existed since the Second World War where the super powers allowed direct confrontations only in far away and often irrelevant places (Vietnam, Afghanistan and Angola) but were able to avoid a frontal collision.

The violent conflict in what used to be Yugoslavia, the more or less peaceful «divorce» in Czechoslovakia, the latent crises in Poland and the complex problems in the Republics of the former Soviet Union illustrate clearly that these system transformation processes are and can be the source of new types and kinds of conflicts for which there will be no easy solution. Unfortunately, the political and economic sciences individually and collectively do not have the necessary theoretical and methodological *instrumentarium* to come up with clear and precise recommendations as to what strategies of change could make this transformation process stable and lead to the common goal, the establishment of a democratic system with a market economy where social justice prevails. Political and economic theory has been more concerned with explaining, either revolutionary or evolutionary change of political and economic systems. What is happening today is a combination of both: on the one hand, a radical change of the economic and political paradigm in Eastern-Europe turning from socialism to the principles of political democracy and market economies; on the other hand, the process of transforming all aspects of the social, political and economic system requires the development of new institutions, the development of new decision processes and the development of new policies – a task which cannot be accomplished overnight, a challenge which requires a truly evolutionary process in both theory and practice.

The inadequacy of economic and political theory to deal with the developments in Eastern Europe has been demonstrated by Wagner (1991).<sup>1</sup> He argues that societal systems can be categorized and subdivided in subsystems, primarily the economic, the political and the socio- cultural system and that crisis in the legitimacy in any one of these subsystems will create a crisis of the total system.

Eastern Europe is faced with a fundamental crisis of the economic and the political system and the re-emergence of old social, cultural, ethnic, religious and national identities. Between the subsystems, spillover effects, reinforcement and feed-back effects can lead to an intensification of the crisis of the whole system and endanger the stability of the transformation process. Consequently, these systems in transition are faced with the difficult task of managing system change in such a way that the inevitable instability of the transformation process can be kept at a minimum and not allowed to endanger its success. The question of how to manage a stable transformation process in subsystems and in the system as a whole is the major issue we intend to focus our attention on this investigation.

Wagner suggests that two fundamental questions have to be answered in this context:

- What were the factors that caused the change of the system and how can they be explained?
- What is the dynamic process of subsystem interdependency i.e. the dynamics of system change?

We are primarily concerned about the second question and therefore emphasize the factors that have contributed to stable transformation processes in some societies leading to new stable equilibria in and between subsystems.

Successful, less successful and unsuccessful examples of system transformation processes leading to or trying to lead to stable democratic market economies are plentiful. After the death of Franco, Spain<sup>2</sup> managed to cautiously but consistently convert political institutions that were already existing during the Franco regime to typically democratic institutions. It was not a radical revolution but an evolutionary change to democratic processes and institutions that went hand in hand with other substantive changes in the social, political and economic subsystem. Despite several attempts by the military to reverse this process, it was the strong support which the King and the vast majority of the population gave to this process of transformation that made it a success.

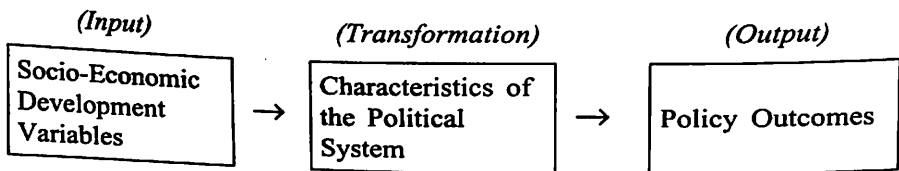
Similar developments have taken place in Portugal and in several Latin American countries<sup>3</sup>. There are also several examples where the transformation process has not been successful, where political systems have not been able to manage the transformation process in such a way that a new stable system and subsystems would emerge. Turkey is an example but also some countries in Latin America<sup>4</sup>. In

the case of South Africa the verdict is not in yet<sup>5</sup>. And of course also in Eastern Europe, despite some hopeful beginnings, it is too early to render final judgement.

These developments pose a challenge to economics and political science to develop theoretical concepts that can explain the dynamics of these system transformation processes and that can suggest strategies for a successful management transition and ensure stability. We propose to take the first steps towards developing such a conceptual framework by interpreting elements of relevant economic and political theory. We shall do this by first taking a critical look at relevant theoretical concepts in political science and economics before, in part III, outlining in a rather tentative and general way one proposal for an integrated approach.

## Part I: The Concept of the Political System and the Policy Approach

The inadequacy of current political theory to explain stable transformation processes and to suggest strategies for a successful management of change, can be illustrated by first looking at policy theory and systems theory in political science. Traditional approaches to this study of political systems treated public policies, the outcomes of the political decision processes, as the major dependent variable with socio-economic development conditions and characteristics of the political system as intervening variable. In a very simplified way this concept can be described as follows:



*Graph 1*

Clearly this model is oversimplified and political scientists have used much more complex concepts (see Fig. 2), but they have generally proceeded under the assumption that certain inputs (demands

and supports) and characteristics of social and economic development are transformed by the political system into policy outputs or public policies. For political scientists the interesting question was whether differences in political system characteristics (transformation process) could explain differences in policy outcomes i.e. the old question whether «politics matters».<sup>6</sup>

The basic elements of this «simple» model are still the same as in Figure 1. There are the inputs, the conversion or transformation mechanisms and the outputs. Added to it are impact or effects. By simply differentiating each of these four categories – for instance, differentiating inputs between system characteristics, political characteristics and problem characteristics, in the political system between policy formation process and structure, and policy implementation process and structure – the complexity is significantly increased. By introducing a differentiation in outputs between target reaction and system reaction which will lead to the eventual outcomes, complexity is further increased. Finally, by specifying the feedback loop, the policy-making system looks, at least on first sight, rather confusing. In this conception, relationships have been designated by arrows, they are not quantified – only their general direction is given. If these would be additionally included, the picture would become almost incomprehensible.

A vast literature developed on this subject matter since the 1960s in the social sciences but particularly in political science. Ever new variables, and ever new measures were used to analyze the relationship between social economic development variables, the characteristics of the political systems and policy outputs or outcomes.<sup>7</sup>

Three issues in this literature are of significance for our concern. They represent a major preoccupation of scholars who worked in this area. They are:

- An analysis of the policy process itself, i.e. how policies are made and at a later stage also how they are applied and implemented.
- The development of a typology of the major types of policies that modern states use;
- The examination of the instruments which are used to achieve the objectives of public policies.

Prior to taking a brief look at the theoretical concepts used in these three areas of policy analysis, it may be worthwhile to think briefly about what «policy» actually means. The often quoted phrase from Harry Eckstein «In political science the safe bet usually is,

that even widely used concepts are not widely understood in a uniform, unambiguous manner» applies here.<sup>8</sup>

Among the dozens of definitions that can be found in this body of literature the most useful for our purposes is that of Hecló who defines policy as «something 'bigger' than particular decisions, but 'smaller' than general social movements... at its core, policy is a course of action intended to accomplish some end».<sup>9</sup> This definition of policy suggests two important dimensions. First, policy is more than small individual political decisions, secondly, that it is a program of actions that tries to accomplish certain social, political or economic objectives. Although this definition maybe criticized or maybe unsatisfactory to some, it suffices for our purpose to narrow down our concern and specify the general field of our concerns.

### 1.1. The Policy Process

Focussing on policy and the process of policy-making marked in political science a turnaway from the prevalent legalistic and institutional orientation. Viewing politics as an intricate process of interactions between individuals and groups brought a new dynamic dimension to political analysis. The key concern were no longer such concepts as state, parliament, executive, parties, but rather actors and decision-makers who worked in a particular institutional context which channeled their actions and activities, but whose personal role and interactions with other actors were more important than the formal institutional context. Institutions were not looked upon as institutions, but as elements in the policy process according to the functions they played in this process.

This orientation in political science certainly brought new insights, but it also lead to new problems. The first problem encountered was the differentiation between the policy process and the environment (social, economic, political) in which it took place and to define this relationship. The concept «political system» was used to describe the general context in which policy processes took place. But where would one have to draw the borders of this system, particularly with respect to the «social» and «economic» system? Were «social» and «economic» actors, i.e. interest groups, associations, large enterprises, part of the political system or were they part of the environment in which the «political» policy process was taking place. The lines between the environment and the policy system proved to be very difficult to draw both theoretically and much mo-



re difficult operationally. Moreover, it was realized that a great number of actors played different roles in different situations; they had to be counted as part of the system in certain aspects of their actions and as part of the environment in others. Conceptualizing the policy process in general terms as it was done by Easton and in our own little graph on page 5, above, was relatively easy. Trying to define it, and operationalize it for research purposes proved much more difficult, and unfortunately, many scholars circumvented the problem by using –easily available– highly aggregated data and by treating the political system as a black box which transformed– somehow miraculously «input» into «output». This approach offered researchers two advantages: first, they could avoid facing the institutional issue –after all «institutions» had come out of fashion– and second, it enabled them to apply advanced statistical methods and thus gave their work the appearance of being truly «scientific».

Efforts to develop more adequate concepts of the policy processes inevitably led to a very complex reconstruction of an even more complex social process. Over time concepts of the policy process became ever more detailed and ever more complex. Figure 3 shows what the author called a «simple model» of the policy process.

Starting at the top of the cycle we begin with the policy development stage. In this stage the European Commission develops a proposal for a Legal Act. Let us say in this case, a directive. It is followed by a policy decision; in our case the Council decides to pass –usually a changed version of– this proposal. The output is a directive. In the next stage Member States are obliged to transpose the directive into the law of their countries, i.e., transpose it into national law. This will be controlled by the Commission. A feedback loop shows that according to Article 169 (EC Treaty) the Commission has to check whether the transposition has been timely and adequate. After full legal transposition Member States have to pass secondary regulations in order to make the directive applicable in practice. The Commission should –and in most cases does– check this stage in accordance with Art. 169 (also called infringement procedure). National Courts can also check the legality of the application process. The Commission also is supposed to monitor the application of national laws and evaluate this process. If problems arise, the Commission –often on requests from Member States– possibly will begin to redesign and update policy which will result in a new policy proposal to the Council and a new, or an amended directive.

An important distinction in these conceptualizations is the distinction between policy formulation and policy implementation. In the early phase of the literature the focus was primarily the process of policy formulation, i.e. the way political decisions were made, the actors that were involved, the influence they had and what the output of the policy system was. It was later realized that policy could not be fully explained without including also the stage after a policy decision had been made, when it had to be applied and implemented. Moreover, it became clear that in the process of implementation new policy choices had to be faced and that policy was not a finished «product» but that it was something that experienced change and was subject to pressures and counter-pressures in the process of implementation as well.<sup>10</sup>

The policy process literature was confronted with another difficulty: The basic structures of the concepts were linear. It was a sequence of events that were captured in the conceptual framework. Yet, policy clearly was a continuous, dynamic process. Today's political decisions are implemented tomorrow and will give rise to the necessity to make a new decision or adjust the previous decision the day after tomorrow. This dynamic aspect was not captured in the conceptual framework. Although there were always references to feedback loops; in terms of operationalization and in terms of theory these feedback lines were drawn into the graphs but not taken seriously.

Early critics of the simplistic policy approach already had pointed to this problem.<sup>11</sup> The obvious solution was the concept of the policy cycle, where policy decision and implementation processes were not viewed as a linear sequence with possible feedback lines but as a cycle which fed into itself.<sup>12</sup> In its most simple form the policy cycle places the major phases of the policy decision implementation process in a cycle where one stage leads into the other. Policy decision making leads to policy implementation, to policy evaluation, to policy redesign and to policy reformulation or redecision (Fig. 3 gives an example for such a cycle).

The major dependent variable is a stable democratic market system characterized by stable, economic, political and socio-cultural institutions.

How this democratic system evolves and whether it evolves depends on a set of very complex interactions between three major types of variables.

- The existing *system ecology* composed of economic factors such as GDP per capita, infrastructure, political factors such as party system, the quality and competence of the existing administration and socio-cultural factors such as the social security system, education, religion, etc.
- the characteristics of the *system in transition*, composed of economic, political and socio-cultural factors and
- *policy output choices*, key policy choices in the economic, political and socio-cultural areas.

It is our argument that the most important determinant of whether and what kind of stable democratic system will emerge depends on the kind of policy choices that are the result of existing system ecology primarily and the characteristics of the political system in transition. How these factors interact, their relevance and importance, their weight and their consequences can only be answered through empirical research.

This dynamic concept provides a more systematic perspective of the policy process linking the various elements and forcing researchers to include a greater number of variables in their examinations. Policy cycles are different for different policy areas, different policy fields and different political systems. The model pictured in Fig. 3 has been designed to capture the process of European environmental policy. It shows many crucial linkages between different actors on different levels of governments.

The policy analysis literature in political science has been a revitalizing and refreshing innovation to traditional political analysis. As part of the behavioral movement, it brought new ideas, new concepts and, above all, a new concern with empirical research. It saw itself in strong contrast with the disciplines' traditional focus on institutions and legal aspects. It is understandable therefore, that it ignored the linkages between the policy process and political institutions and underestimated the importance and the restraining force of political institutional structures. Nonetheless, it provided a conceptual framework for a large number of very interesting empirical inquiries of high scientific quality. Comparative policy analysis moreover, suggested that there were significant differences between policy processes in different political systems and also with respect to different policy areas. It was necessary therefore, to develop a typology of policy areas from the perspective of a policy process point of view.

## I.2. Typologies of Policy Area

By far the most simple typology is that based on the content of a policy. One can speak therefore of social policy, economic policy, employment policy, etc. But this simple division is not very useful from an analytical point of view, because policy analysis is trying to find commonalities and significant differences across these substantive policy areas. The key question is not to differentiate according to every possible designation of substantive policy areas but to find categories that characterize certain types of policies, that point to significant differences between social policies and economic policies, etc.

Most efforts to arrive at such a typology have been more or less inspired by Theodore Lowi's typology. Principally, Lowi distinguishes only between four types of policies: distributive, regulative, constituent and redistributive.<sup>13</sup> Lowi uses the likelihood of the utilization of coercion in the application of policy. The classical theory scheme is represented in Figure 4.

The primary concern of Lowi was to relate policies to the way they are made and applied. In a sense, he was interested in the relationship between policy and politics, whereby politics for him was the dependent variable.

What has fascinated scholars with Lowi's classification scheme is not this initial concern but rather the use of this typology as a simple descriptive classification of policies. Indeed, it seemed logical to speak of

- distributive policies, where things are being given away;
- regulative policies, where restrictions, rules of conduct and sanctions are stipulated;
- redistributive policies, where resources are taken from one group of people and given to another
- constituent policies as general rules about how decisions are made, rules about the political system itself, rules about power and how it is handled.

Moreover, some of these terms have been used in other fields of inquiry, like economics which suggested that there were indeed fundamental differences between these types of policies with respect to the policy content and process. Unfortunately, the typology is lacking precision, it is rather general including a broad range of phenomena in one category which makes it rather difficult to apply and operationalize the typology in empirical research.

Because of these difficulties, it is not surprising that others have developed alternative typologies most of which are based on the four-fold scheme of Lowi. The most interesting is that of Salisbury and Heinz.<sup>14</sup> The criteria for distinction are not any longer the use of force or coercion, but the characteristics of the decisional system and the demand pattern. Salisbury differentiated between integrated or fragmented political decision system and demand pattern (see Fig. 5). He arrives at the same four categories or types of policy, whereby constitutional policies are considered only a special case of regulative policies. In its place he uses a new type which he calls «self-regulative policy» referring to policy areas where regulation is left to those concerned themselves since the decision system is fragmented and not capable of enforcing a redistributive policy in this case. The further refinement can be found in the Salisbury/Heinz publication of 1970 where the decisional system is replaced by the concept of the «costs» involved in reaching decisions. It is not more than an interesting variation to the same theme.

Also based on Lowi's four-fold typology, Paul Sabatier tried to move further away from it by focussing more on the very content of different types of policies. He differentiates between policies which provide or allocate goods and services and policies which regulate behavior. The other dimension involves the benefits and costs of policies where he distinguishes between zero-sum policies (where clearly one group gains what another group loses) and non-zero-sum policies (where both or all participants can gain from the policy).<sup>15</sup>

It is only a small step from this typology to a typology which is more concerned with instruments of policy, i.e. the kinds of measures taken to realize certain objectives. A typology that stands somewhere between the Sabatier classification and a purely instrumental one, is the one developed by Jann which includes both the Sabatier type categories and adds typical policy instruments such as raising taxes, introducing new «programs» or «spending programs», etc.

Prior to moving to the discussion of the instrument question, we should keep in mind that particularly the Salisbury typology could potentially be very useful in working or applying or developing concepts and frameworks for analysis of political systems in transition, since one of his criteria is the nature of the political system, whether it is integrated and fragmented, although the categories are probably much too general to be directly applicable.

As interesting as these classificatory schemes of the substance of policy are, they all avoid the crucial question of the consequences

of policy content for the stabilization of political systems. They do not ask whether some type of policy is more conducive towards stabilizing institutions and others are less likely to achieve this objective. Nonetheless, some of these categories may be useful in developing our own classificatory scheme of the substance of policy.

### I.3. Instruments of Public Policy

Interestingly enough, one finds little reference to the question of policy instruments in the American literature. It is a domain which was developed primarily in Europe, above all in Germany and in the Scandinavian countries. Perhaps, this is a result partially of a stronger influence of economic theory on policy research in these countries. Perhaps, it is also a consequence of the more intensive concern in the European policy literature with concrete «programs» of public policy ranging from social programs, environmental programs to planning etc. This concern is particularly characteristic of the implementation literature in Europe.<sup>16</sup>

Although the literature is full of very detailed and highly differentiated classification schemes of policy instruments, for our purposes, the most commonly used three-fold differentiation will be sufficient:

- administrative instruments, that is rules and regulations enacted, enforced and controlled by government;
- economic instruments, the use of taxes or other levies or subsidies and incentives to achieve certain objectives;
- informational and consultative instruments, where policy provides guidance and support for citizens, e.g. employment offices, consumer offices, etc.

Basically, this typology of policy instruments distinguishes between the resources that are used in policy application and implementation, i.e. power, money, and information.

These instruments can be further refined and subdivided. Economists have developed a full array of economic instruments. In political science and in public administration a great number of subdivisions have been developed with respect to political-administrative instruments and informational instruments, where also sociology has made some contributions.

In conclusion, it also should be emphasized that specific policies or political programs may employ a whole range of these instru-

ments and their effectiveness and success often depends on a specific mix of political-administrative, economic and informational policy instruments.

#### I.4. An Intermediate Conclusion

Policy analysis in political science is primarily concerned with answering the question of how things are done, how policies are made, how they are implemented, how they are applied. It has only in a few exceptions focused on another important question, namely on what is policy or what should policy be, i.e. the material dimension of policy. This was left by and large to the specialists from other fields, to the economists, to the educational people, to the social policy people, etc. At the same time policy analysis would be an ideal area where the process and the structure oriented scholars would work together with those interested in the what question, the material dimension.

Another weakness of the policy literature has been identified as its failure to link up with institutions, with the structures in which policies take place, in which policies are decided and are implemented. It is also reflected in the very limited discussion of policy instruments, where particularly the relationship between material content, objectives and instruments, involving an economic, political scientist and public management perspective, is virtually absent. If, as we are trying to do, one wants to develop a conceptual framework for political systems in transition, for systems in search of stable democratic institutions, we need to provide these linkages. We need to integrate economic and political science theory, we need to integrate the question of the substance of policy, its relationship to policy process and the actors involved in it and to the institutions that provide the framework for policy making and implementation. In most stable systems these are given and analysts need not ask these questions. Change is taking place slowly, institutional consequences of policy choices and policy processes can be at best observed over long periods of time. In systems in transition the opposite is the case; it is a crucial question and time is a scarce commodity.

This is particularly urgent when the timetable of accession to the European Union dictates the agenda of domestic policy. In fact, there is an umbrella effect, since even those states not now being active candidates but preparing to become candidates, anticipate the requirement to absorb the *acquis communautaire*. In addition, the um-

brella also extends to their traditional trading partners who absorb the acquis, more or less successfully, in order to be associated. The role of the Council of Europe in this should not be overlooked either. Witness the recent decision of the Republic of Romania to abolish the death penalty (April 2003).

## Part II: The transition to a social market economy and a democratic political institution as an economic problem

The sudden demise of the state socialist economies in Eastern Europe has caught most professional economists in the West by surprise. Within months, however, a stream of literature began to flow and economists of all kinds of persuasions tried to apply their analytical tools to the new phenomenon. While the profession functioned effectively in addressing a newly urgent problem, it did not in systematically building on the extant work on economic systems that had accompanied the rise and development of the market economies for about the middle of this century.

Three instances highlight this point.

1. János Kornai has certainly established himself as one of the leading authorities in the economics of transitions from one system to the other. In a nutshell, his *The Road to a Free Economy* (1990)<sup>17</sup> can be summarized in the sequence of steps outlined in his pages 39-40:
  1. Freedom to establish a firm; free entry into the production sphere.
  2. Free prices, based on a free contract between the buyer and the seller.
  3. Unrestricted right to rent out privately owned assets, again on the basis of a free contract between the lessor and the lessee. Among other things, these transactions should include the free renting out of privately owned apartments or real estate, with the rental freely agreed upon between the lessor and the lessee.
  4. Unrestricted right to employ people in all cases where the employer belongs to the private sector household or privately owned firm. The employer and the employee must be free to agree on wages.



5. Unrestricted right to accumulate, sell, or buy any article of value (e.g., rare metal).
6. Unrestricted right to accumulate, sell, or buy foreign currency through transactions within the private sector and between Hungarian and foreign citizens.
7. Unrestricted right to take out and bring in domestic and foreign currencies.
8. Free foreign trade activity, in which the member of the private sector has the unrestricted right to export and import.
9. Unrestricted right to lend money, with credit terms freely agreed upon between the creditor and the debtor.
10. Freedom of financial investment in the private ventures of other individuals.
11. Freedom to sell and buy, at free prices, any privately owned apartment, real estate, or other asset.

For a mathematical proof, the sequence in which particular conditions are ordered makes no difference. For a process of transition, the sequence does indeed make a difference. It is obvious that the sequence indicated by Kornai is by no means an intended sequence. It is simply a random compilation of different conditions a particular system has to respond to. Alternatively, one could think of the preconditions to be met for a market economy to evolve in a step-wise manner.

There is still no generally agreed upon theory of the transformation of economic systems from central planning in physical units to decentral market coordination. Yet there is agreement that a number of institutional prerequisites are necessary for the operation of a market economy; and additional prerequisite institutions are needed if the market economy towards which transformation is proceeding is to be embedded into a social welfare state. Both political and economic considerations will determine the best possible path of the transformation and, in particular, the sequence of institutional reform measures as they will have to be implemented one after and building on the other.

The simplest and most basic transformation scenario might consist of at least the following six steps:

1. Establish a property rights regime supported by the three basic guarantees of private property, freedom of contract, and liability.
2. Define a transformation path in a politically binding way.
3. Facilitate the introduction of a free monetary system by granting

- citizens the freedom to choose whichever currency or other medium of exchange they prefer for their own transactions.
4. Define in a legally binding way the assignment of property rights in the economy and society in order to provide a starting point for the transformation path.
  5. Establish a financially sound and independent system of social security which is kept separate from the state budget; as a funding source, this system could receive the exclusive right to levy a sales or value added tax.
  6. Stabilize the public finances.

This is an example of how a sequence could be developed that builds on the institutional preconditions to go from step  $n$  to step  $n+1$ .

2. A second example may be drawn from an outstanding journal now primarily devoted to economic change in Eastern Europe. Alfred Steinherr's «Essential Ingredients for Reform in Eastern Europe» surprises the reader with statements like these: «Because flexible exchange rates generate too much instability and exert too little constraint on domestic policies, the exchange rate must be fixed (most usefully in terms of the ECU). Convertibility may then remain limited to current account transactions (p.5). And again, «Radical reforms have the intuitive appeal of not rocking the boat and tackling problems measure for measure in a logical order. Experience shows, however, that such a logical ordering is not possible in that, during the transition from one system to another, contradictions developed which may deteriorate performance even further and reinforce conservative forces.» (p. 7 – a reference to the experience of partial price reform in Hungary follows in order to collaborate the point). Note that his statement in favor of the big bang approach rests on the unproved assumption that there is no logical order to the sequence of reform steps.
3. The third example is a bibliography kindly shared with me by an official from the Deutsche Bundesbank in Frankfurt. The bibliography is entitled (in my translation): Transformation of the Economic Order of the Former DDR: Economic restructuring in Eastern Germany: a Guide to the Literature. The part super-scribed «basic literature on the transformation of economic systems» running from pages 13 through page 28 contains none of the leading works in German language economic theory on the subject.

Let me give a few hints to this literature in the second part of this appendix.

1. Although the pre-Smithian Cameralist literature already established a number of important criteria to be met by successful plans for transforming agricultural into market economies (to the benefit of the Cameralist' state), the Smithian inquiry into the nature and *causes* of the Wealth of Nations set of a new wave on inquiry in the continental countries. They were faced with different conditions, since sea faring trade and the conquering of natural resources and the establishment of monopolies of international trade were not at their disposal<sup>18</sup>. In the German language countries, the Smithian heritage had to be translated into different institutional structures and consequently into different political responses or programs. Most challenging was the task met by the German professors in Russia, since here, they were unable to use fairly stressing reform instruments while facing a particularly underdeveloped economy. This clash of high-powered knowledge with arcane economic processes may serve as a backdrop for our contemporarily political and academic challenge<sup>19</sup>. The leading authority in this group is obviously Heinrich von Storch, who combined his economic reform strategy, taken up from Smith, but adapted it to the Russian circumstances with a theory of civilization, or a theory of the political circumstances under which, under the conditions of his Russian environment, a market economy might have a chance to evolve.
2. The second great classic in the theory of transitions from system to system is the man who actually introduced the term «economic systems» into the discipline of economics. Werner Sombart in his monumental *Modern Capitalism*<sup>20</sup> carefully traced the development from precapitalism to early capitalism, mature capitalism, high capitalism and late (regulated) capitalism by indicating precisely the institutional conditions at the path leading from system to the other. However, he did not systematically introduce the political element; following Pareto's advise, it seems he shied back from a systematic integration of the political element, which he left to his friend Max Weber. Pareto, on the other hand, in his «Les Systemes Socialistes» and later in his «Trattato di Sociologia generale» gives a number of conditions for successful transitions from one system to the other.

3. The archetypal researcher into transition processes from one economic system to the other was, of course, Gustav Schmoller, whose concept of economics involved the emphasis on the socio-economic conditions for particular economic systems to develop. He did not shy away from designing political economic systems (such as the Modern Welfare State with the three pillars of social insurance) if he could, by way of historical or comparative research, establish viable institutions that had served similar purposes in similar societies. His *Grundriss der Allgemeinen Volkswirtschaftslehre* is replete with catalogues of conditions for the various steps of transition paths from one stage of socio-economic development to the other.

## Conclusion

Our hypothesis is that the substance of specific social and economic policies will shape if not determine the type of new institutions in the three subsystems that will evolve in a transformation process. We are assuming that there are some policies that are «better» than others in helping in the establishment of democratic institutions, of course depending on crucial existing conditions in the system (system ecology). If it is possible to identify «successful» policies and how they are related to institutional change in the desired direction, it should be possible to develop a strategy for managing a stable system transformation. The elements for such a conceptual framework are presented –again highly oversimplified– in graph 6.

The major shortcoming of graph 6 lies in the fact that the graph does not help in elucidating the dynamic structure of the transitional process. As we emphasized earlier, the main weakness of the naive literature on transitional processes lies in the fact that conditions are listed without giving the logical sequence of their introduction. We correct this weakness by indicating one conceivable scenario by way of example. The scenario indicated is shocking and baffling to most practitioners, because it flies in the face of textbook wisdom. Yet, this textbook wisdom is irrelevant, as it does not address systems in transition. E.g. one of the participants in the conference in Bad Homburg, where the papers combined in this volume were discussed, was upset by the proposal to fund the social security system with the proceeds of a sales tax and thereby wresting this po-

werful tax from government, probably with the result of reducing the government's ability to lay a credible claim on resources in that economy.

Every transitional process that we have so far studied not only shows us that the textbook conditions were never met even in the most successful cases. The most important results not addressed in the literature we have so far been able to survey is the fact that transitional processes need transitional institutions which will be created and then disappear. The Treuhand Anstalt is just one and very specific an example for such a transitional institution. More specifically, we hypothesize that a successful transformation strategy will involve the sequential creation of institutions which by their own design can not survive.

The basic rationale of this hypothesis is by no means new. A prototype can be found in Friedrich List's<sup>21</sup> *Developmental Strategy of Industrialization*. The often-misunderstood argument for a protective tariff was carefully drafted in such a way, that the tariff would have to be revoked in the case of both possible outcomes. If the protected industry did not develop according to plans, the tariff had to be revoked as it was too expensive in terms of tariff expenditures or revenue lost. In case it succeeded, the tariff would have to be repealed by virtue of its insignificance from a revenue point of view, and most likely also because it would prove to be a political liability. This variant of constitutional political economy, which emphasizes the design of self-enforcing institutions which as their best strategy have an incentive to complete the task assigned and then devolve, has not been emphasized much after List pioneered it. It is here where much more work will remain to be done, once enough reliable information is available about successful transformation processes.

## Bibliography

It is common practice in English publishing to insist on a one to one relationship between one of a piece of literature and one source and vice versa. While we are always glad to give credit to our colleagues and the sources we have been allowed to draw upon, in this case we face an embarrassment of richness. The discussions that led to this corporate piece were informed by what we know ourselves

through travel and extensive consulting and the references we cite. But not every item in the list of references resurfaces in the list. In this ..., we insist on couraging the entire referee list, without a one to one relationship; we simply cannot properly credit some of our ideas, but we have an idea about the common stock of knowledge on which we draw.

## Notes

1. Helmuth Wagner, «Einige Theorien des Systemwandels im Vergleich - und Ihre Anwendbarkeit für die Erklärung des gegenwärtigen Reformprozesses in Osteuropa». In: Jürgen Backhaus (ed.), *Systemwandel und Reform in östlichen Wirtschaften*. Marburg: Metropolis, 1991, pp. 17-40.

2. See: Jose Casas Pardo and Giuseppe Eusepi, «Individualism vs. Idealism: Constitutional and Institutional Reform in the Nineties», in *Torino: European Public Choice Society*, 1992; and José Casas Pardo and Jack Wiseman, «The Political Economy of the Spanish 1978 Constitution Making Process», in *Beaune: European Public Choice Society*, 1991. Also: José Casas Pardo and Friedrich Schneider, *Current Issues in Public Choice*, Cheltenham, Gloss: Edward Elgar Publishing, 1997.

3. Terry Lynn Karl, «Dilemmas of Democratization in Latin America», *Comparative Politics*, Vol. 23, Oct. 1990, pp. 1-22 and Guillermo O' Donnell, Philippe Schnatler, Laurence Whitehead, (eds.), *Transition from Authoritarian Rule*, Baltimore, John Hopkins Press, 1986.

4. Guillermo O'Donnell, Philippe Schnitter, *Tentative Conclusions about Uncertain Transition*, Baltimore, John Hopkins Press, 1986; Metin Heper and Ahmet Erin (eds.), *State, Democracy and the Military, Turkey in the 1980's*, New York, de Gruyter, 1988, (especially Part V); Metin Heper, «The Executive in the Third Turkish Republic, 1982-'89» *Governance*, Vol. 3, Nr.3, 1990, pp. 299-319.

5. Fanie Cloete, Lawrence Schlemmer and Daan van Vuuren, eds., *Policy Options for a New South Africa*, Pretoria, Human Sciences Research Council, 1991.

6. The basic concept underlying the literature was first proposed by David Easton in his already classic book *The Political System*, New York, Knopf, 1953. He further developed it in a later article «An Approach to the Analysis of Political Systems», *World Politics*, Vol. 9, 1957, pp. 383-400 and in *A Systems Analysis of Political Life*, New York, Wiley, 1965. Following Eastons pioneering work, a virtual explosion of literature occurred during the 1960s in the United States which tried to apply and submit this concept to quantitative tests. The first of these was an article by Richard E. Dawson and James A. Robinson, «Interparty Competition, Economic Varia-

bles and Welfare Policies in the American States», *Journal of Politics*, Vol. 25 (May 1963), pp. 265-289. Subsequent contributions have been made by three political scientists: Thomas Dye, Ira Sharkansky and Richard Hofferbert. See Thomas Dye, *Politics, Economics and the Public: Policy Outcomes in the American States*, Chicago Rand McNally, 1966; Ira Sharkansky, *Spending in the American States*, Chicago Rand McNally, 1968; idem, *The Politics of Taxing and Spending*, Indianapolis Bobbs-Merrill, 1969; idem, *Regionalism in American Politics*, Indianapolis Bobbs-Merrill, 1970; idem, *Policy Analysis in Political Science*, Chicago Markham 1970; as well as a number of articles whose substance mostly reappears in these books; and Richard I. Hofferbert, «Socio-Economic Dimensions of the American States: 1890-1960», *Midwest Journal of Politics*, Vol. 12 August 1968, pp. 401-418, and idem «The Relation Between Public Policy and Some Structural and Environmental Variables in the American States», *American Political Science Review*, Vol. 60 March 1966, pp. 73-82.

A first critical analysis of the conceptual framework underlying this literature can be found in Stuart H. Rakoff and Guenther F. Schaefer, «Politics, Policy and Political Science, Theoretical Alternatives», *Politics and Society*, 1970, vol. 1, pp. 51-77. One of the most convincing answers to these early findings which were duplicated later in the 1970s countless times is L.J. Sharpe and K. Newton, *Does Politics Matter? The Determinance of Public Policy*, Oxford, The Clarendon Press, 1984. Sharp and Newton demonstrate using essentially the basic assumption of this approach that intelligent use of empirical data and particularly the appropriate disaggregation of standard statistical information can show a definite linkage between political variables and policy outputs.

7. A good old view of these literatures provided by Thomas R. Dye, *Understanding Public Policy*, Englewood and Cliffs, New Jersey, 1975 and *Policy Analysis, What Governments Do, Why they Do it and what Difference it Makes*, Birmingham, Alabama, 1976. Richard I. Hofferbert, *The Study of Public Policy*, New York, Indianapolis, 1974. Werner Jann, *Kategorien der Policy-Forschung*, Speyrer Arbeitshefte, 37 (1981) and Adrienne Windhoff-Heritier, *Policy Analyse, eine Einführung*, Frankfurt, 1987.

8. Harry Eckstein, «A Case Study and Theory in Political Science», in Fred I. Greenstein and Norman W. Polsby, (ed.), *Handbook of Political Science*, Vol. 7, Redding, Massachusetts 1975.

9. Hugh Heclo, «Policy Analysis», *British Journal of Political Sciences*, 1972, p. 84.

10. A survey of the last literature on policy implementation see Daniel A. Mazmanian and Paul A. Sabatier, (eds.), *Effective Policy Implementation*, Lexington, Massachusetts, 1981, and Renate Mayntz, (ed.), *Implementation politischer Programme*, Königstein, Taunus, 1980, and *Implementation politischer Programme II*, Opladen, 1983.

11. Rakoff and Schaefer, *op. cit.*

12. Judith V. May and Aaron Wildavsky, *The Policy Cycle*, London, 1978; and W. Jann, *op. cit.*

13. Theodore J. Lowi, «Four Systems of Policy, Politics and Choice», in *Public Administration Review*, 1972, pp. 298-310, contains the most elaborate description of this classification. It can also be found already in an earlier publication in «American Business, Public Policy, Case Studies and Political Theory», in *World Politics*, vol. 16, 1964, pp. 677-715. Some modifications are found in «Public Policy and Bureaucracy in the United States and in France», in Douglas Ashford, (ed.), *Comparing Public Policies*, Amsterdam, 1978, (Sage Yearbook of Politics and Public Policy).

14. Robert Salisbury «The Analysis of Public Policy», in Austin Ranney, (ed.), *Political Science and Public Policy*, Chicago, 1968, pp. 151-157 and Robert A. Salisbury and John Heinz «A Theory of Policy Analysis and Some Preliminary Applications», in Ira Sharkansky, (ed.), *Policy Analysis in Political Science*, Chicago, 1970, pp. 39-60.

15. Paul Sabatier, «Social Movements and Regulatory Agencies: Toward a more Adequate - and less Pessimistic - Theory of 'Clientele Capture'», in *Policy Sciences*, vol. 6, 1975, pp. 301-327, and Paul Sabatier and Daniel Mazmanian, «The Implementation of Public Policy - A Framework of Analysis», in *Policy Studies*, 1980, pp. 538-560.

16. Some of the key publications in this area are Renate Mayntz «Die Implementation politischer Programme: Theoretische Überlegungen zu einem neuen Forschungsgebiet», in: *Die Verwaltung*, No. 1, 1977, pp. 51-78, also the 2 volumes on implementation research edited by Renate Mayntz, cited in note 10. In addition, Claus Offe, *Berufsbildungsreform, eine Fall-studie über Reformpolitik*, Frankfurt, 1975 and Fritz W. Scharpf, «Theorie der Politikverflechtung», in Scharpf, Reissert, Schnabel, *Politikverflechtung: Theorie und Empirie des Kooperativen Föderalismus in der Bundesrepublik*, Kronberg, 1976. An overview is provided by Werner Jann and Eva Kronenwett «Handlungsspielräume und Entscheidungsfähigkeit des Staates untersucht am Beispiel der Implementation politischer Programme, Empirisches Untersuchungsraaster», Speyer, 1978, in *Speyerer Arbeitshefte*, No. 27 and Jann, *op. cit.*, pp. 60-69.

17. János Kornai, *The Road to a Free Economy: Shifting from a socialist system again: the Example of Hungary*, New York, WW Norton, 1990.

18. For a survey of this literature please refer to Jürgen Backhaus and Richard E. Wagner, *The Cameralist's Public Choice Perspective*, Public Choice, 53, 1987, pp. 3-20

19. Jürgen G. Backhaus, «Die West-östliche Migration Deutscher Ökonomen: Ein neu zu Schreibendes Kapitel in der Geschichte Der Wirtschaftswissenschaften», Vortrag anlässlich der Tagung des Dogmenhistorischen Ausschusses des Vereins für Socialpolitik in der Kartause Ittingen vom 6. bis 8. Oktober 1991.

20. Review article by Jürgen Backhaus: Werner Sombart: «MODERN CA-



PITALISM». *Kyklos*, 42.4, 1989, pp. 599-611. Reprinted in: Mark Blaug (ed.) *Pioneers in Economics Section III*. Aldershot: Edward Elgar Publishing Ltd.

21. For a more detailed exposition, see Jurgen Backhaus, «Friedrich List and the Political Economy of Protective Tariffs», in: S. Todd Lowry (ed.), *Perspectives on the History of Economic Thought VII*, Cheltenham: Edward Elgar Publishing Limited, 1992, pp. 142-156.

## References

- Abouchar, A. (1990), «The Soviet Economy: Whence, Where, Whither?, Why?», paper presented at the Conference of Allied Science Associations, Washington, D.C., December.
- Aeppl, T. and J. Mayer (1990), «Reckoning Day - East Germany Is Set To Make Great Leap, But It Ponders Costs», *The Wall Street Journal Europe*, March, 16-17;1,2.
- Alyushin, A.L. (1992), «The Paternalistic Tradition and Russia's Transition to Liberal Democracy», *Democratic Institutions 1*, Carnegie Council on Ethics and International Affairs and The Center for Research and Communication.
- Backhaus, J. and Wagner R. (1987), *The Cameralist's Public Choice Perspective*, *Public Choice*, 53, pp. 3-20.
- Backhaus, J. (1989), «Werner Sombart: MODERN CAPITALISM», *Kyklos*, 42.4, pp. 599-611, reprinted in Blaug, M. (ed.), *Pioneers in Economics Section III*, Aldershot: Edward Elgar Publishing Ltd.
- Backhaus, J. (1991), «Die West-östliche Migration Deutscher Ökonomen: Ein neu zu schreibendes Kapitel in der Geschichte der Wirtschaftswissenschaften», vortrag anlässlich der Tagung des Dogmenhistorischen Ausschusses des Vereins für Socialpolitik in der Kartause Ittingen, 6-8 Okt.
- Backhaus, J. (1992), «Friedrich List and the Political Economy of Protective Tariffs», in Lowry, S.T. (ed.), *Perspectives on the History of Economic Thought VII*, Cheltenham: Edward Elgar Publishing Ltd., pp. 142-156.
- Backhaus J., Tchipev, P. and Stephen, F. (eds.), «Mass Privatisation Schemes in Central and East European Countries. Implications on Corporate Governance», Sofia: GorexPress, 1998
- Balabkins, N.W. (1990), Book Review of Kornai, J. (1990), *The road to a free economy: shifting from a socialist system. The example of Hungary*. New York: Norton & Company in *Journal of Institutional and Theoretical Economics*, September 12.
- Bates, R.H. (1991), «The Economics of Transitions to Democracy.» *Political Science & Politics*, March 24-27, 24-27.

- Bernholz, P. (1992), «On the Political Economy of the Transformation of Political and Economic Regimes.» Paper prepared for the Eighth Annual Lecture in the Virginia Political Economy Lecture Series, Fairfax: Center for Study of Public Choice, March.
- Betz, H.K. (1991), Book Review of «Not by Theory alone: The Economics of Gustav von Schmoller and its Legacy to America» by Nicholas W. Balabkins, *History of Political Economy* 3, 539-541.
- Bohnet, A. and M. Reichhardt (1993), «Der Beitrag der Transaktionskostenökonomik zu einer Theorie der Transformation von Wirtschaftsordnungen, (The Contribution of Transaction Cost Economics to a Theory of Economic Transformation.)» *Jahrbücher für Nationalökonomie und Statistik* 212 (3-4), 204-226.
- Bos, D. (1991), «Privatization and the Transition from Planned to Market Economies: Some Thoughts About Germany 1991», *Annals of Public and Cooperative Economics* 62, 183-194.
- Brennan, G. and J. Casas Pardo (1991), «A Reading of the Spanish Constitution», *Constitutional Political Economy* 2(1), 53-79.
- Brown, A.A. (1990), «A Grants Economics Perspective on Kornai's Proposal for the Transformation from a Socialist Economic System», paper presented at the 1990 Meetings of the Allied Social Sciences Associations, Washington, D.C., December 29.
- Buchanan, J.M. (1990), «An American Perspective on Europe's Constitutional Opportunity», paper presented at the Mt. Pelerin Society General Meeting, Munich, Germany, September.
- Buck, H.F. and H.G. Bauer (ed.) (1991), *Transformation der Wirtschaftsordnung der ehemaligen DDR. Wirtschaftliche Erneuerung in Ostdeutschland - Literaturführer.* (Literatureguide: Transformation of the Economic Order of the former GDR, Economic Renewal in East Germany), Bonn: Gesamtdeutsches Institut.
- Casas, J. Pardo and Eusepi, G. (1992), «Individualism vs. Idealism: Constitutional and Institutional Reform in the Nineties», in *Torino: European Public Choice Society*
- Casas, J. Pardo and Wisseman, J. (1991), «The Political Economy of the Spanish 1978 Constitution Making Process», in *Beaune: European Public Choice Society*.
- Casas, J. Pardo, and Schneider, F. (1997), *Current Issues in Public Choice*, Cheltenham, Gloss: Edward Elgar Publishing.
- Chaloupek, G. (1992), «The Concept of Maturity and the Transformation of Economic Systems.» Paper presented to the symposium in honour of Josef Steindl, April.
- Cloete, F., Schlemmer, L and van Vuuren, D. (eds.) (1991), *Policy Options for a New South Africa*, Pretoria, Human Sciences Research Council.
- Dawson, R.E. and Robinson, J.A., (May 1963), «Interparty Competition,

- Economic Variables and Welfare Policies in the American States», *Journal of Politics*, Vol. 25, pp. 265-289.
- Dietrich, S. (1990), «Im grauen Polen grünen die ersten Pflänzchen der Marktwirtschaft.» *Frankfurter Allgemeine Zeitung*, February 10, 3.
- Donges, J.B., W. Engels, W. Hamm, W. Möschel and O. Sievert (ed.) (1990), «Wirtschaftsreformen in der DDR - Das Eigentum in der Marktwirtschaft. (Economic Reforms in the GDR - Property in the Market Economy)», *Frankfurter Institut - Argumente zur Wirtschafts-politik* 31.
- Donges, J.B., W. Engels, W. Hamm, W. Möschel and O. Sievert (ed.) (1991), «Darauf kommt es an beim Aufbau des geeinten Deutschland», *Frankfurter Institut - Argumente zur Wirtschaftspolitik* 38.
- Donges, J.B., W. Engels, W. Hamm, W. Moschel and O. Sievert (ed.) (1992), *Zur Wirtschaftsreform in Osteuropa*, frankfurter Institut für wirtschaftspolitische Forschung 24.
- Dye, T. (1966), *Politics, Economics and the Public: Policy Outcomes in the American States*, Chicago Rand McNally.
- Dye, T. (1975), *Understanding Public Policy*, Englewood and Cliffs, New Jersey.
- Dye, T. (1976), *Policy Analysis, What Governments Do, Why they Do it and what Difference it Makes*, Birmingham, Alabama.
- Easton, D. (1953), *The Political System*, New York: Knopf.
- Easton, D. (1957), «An Approach to the Analysis of Political Systems», *World Politics*, Vol. 9, pp. 383-400
- Easton, D. (1965), *A Systems Analysis of Political Life*, New York: Wiley.
- Eckstein, H. (1975), «A Case Study and Theory in Political Science», in Greenstein, F.I. and Polsby, N.W. (ed.), *Handbook of Political Science*, Vol. 7, Redding, Massachusettes.
- Fender, J. and D. Laing (1992), «The Macroeconomics of a Reforming Centrally Planned Economy with Queuing and Resale», working Paper, department of Economics, the Pennsylvania State University, April.
- Fender, J. and D. Laing (1992), «The Transition from Central Planning: An Intertemporal Macroeconomic Model», working Paper, Department of Economics, The Pennsylvania State University, May.
- Fisher, A. (1990), «Privatisation could help meet the bill for German unification», *Financial Times*, October 24.
- Forman, C. (1989), «Slow Going - A Buyout in Hungary Spotlights the Snags in East Bloc Reforms», *the Wall Street Journal/Europe*, November 29; 1,4.
- Furstenberg, G.M. von (1990), «Reviving East German Property», *The Wall Street Journal/Europe*, March 8.
- General-Anzeiger Bonn (1990), «Bundesbank möchte Einheit durch Privatisierung finanzieren», October 4.

- Gewirtz, C. (1990), «Economic View East: It's Bleak», *International Herald Tribune*, February 28.
- Gress, D.R. (1990), «German-Polish Mutual Interests Count», *the Wall Street Journal/Europe*, March 2-3.
- Gretschmann, K. (1993), «Neue ökonomische Institutionsanalyse», European Institute of Public Administration.
- Haig, M.H. jr (1990), «... Little if Reunification Comes on Soviet Terms», *The Wall Street Journal/Europe*, March 2-3.
- Hagemann, H. (1993), «Makroökonomische Konsequenzen der deutschen Einigung (Macroeconomic Consequences of the German Reunification)», *Probleme der Einheit* 11, Marburg: Metropolis-Verlag.
- Hampe, P. (1990), «Neue Finanzierungsideen für die Einheit», *Süddeutsche Zeitung*, October 11.
- Hauser, K. (1990), Book Review of Balabkins, N.W. (1988), *Not by Theory alone*, Berlin: Duncker & Humblot in *Journal of Institutional and Theoretical Economics* 146(2), 374-377.
- Heck, H. (1990), «Runter mit den Schulden», *Die Welt*, October 5.
- Hecló, H. (1972), *British Journal of Political Sciences*, p. 84
- Hefty, G.P. (1990), «Eine Entscheidung, die alles änderte», *Frankfurter Allgemeine Zeitung*, February 3.
- Helms, R. (1990), «Why Do We All Accept the Rush?», *The Wall Street Journal/Europe*, February 27.
- Helmstadter, E. (1991), «Eigentum und Kapitalwirtschaft in der Ordnungspolitik (Property Rights and the Use of Capital in the Theory of Economic Systems)», *ORDO* 42, 235-251.
- Helmstadter, E. (1991), «Neue Bundesländer als Industriestandort», *Kommentare zur Wirtschaftspolitik*, 14.
- Helmstadter, E. (1991), «Systemführerschaft. Wettbewerbstheoretische Begründung und wettbewerbpolitische Schlußfolgerung», *Kommentare zur Wirtschaftspolitik*, 10.
- Heper, M. and Erin, A. (eds.) (1988), *State, Democracy and the Military, Turkey in the 1980's*, New York, de Gruyter, (especially Part V).
- Heper, M. (1990), «The Executive in the Third Turkish Republic, 1982-'89» *Governance*, Vol. 3, Nr. 3, 1990, pp. 299-319.
- Hofferbert, R.I. (1968), «Socio-Economic Dimensions of the American States: 1890-1960», *Midwest Journal of Politics*, Vol. 12, August, pp. 401-418.
- Hofferbert, R.I. (1966), «The Relation Between Public Policy and Some Structural and Environmental Variables in the American States», *American Political Science Review*, Vol. 60, March, pp. 73-82.
- Hofferbert, R.I. (1974), *The Study of Public Policy*, New York, Indianapolis.
- Holscher, J. (1994), «Money, Privatisation and Private Property: The Example of Central-East Europe», Working Paper of the European Business Management School, University of Wales, Swansea.

- Holscher, J., A. Jacobsen, H. Tomann and H. Weisfeld (ed.) (1994), *Bedingungen ökonomischer Entwicklung in Zentralösteuropa (2)* (Pre-conditions of Economic Development in Central East Europe), Marburg: Metropolis-Verlag.
- Hoffmann, L. (1990), «Wider die ökonomische Vernunft», *Frankfurter Allgemeine Zeitung*, February 10, 15.
- Hoffmann, W. (1990), «Eine radikale Lösung», *Die Zeit*, October 12.
- Jann, W. and Kronenwett E. (1978), «Handlungsspielräume und Entscheidungsfähigkeit des Staates untersucht am Beispiel der Implementation politischer Programme, Empirisches Untersuchungsrastrer», Speyer, in *Speyerer Arbeitshefte*, No. 27.
- Jann, W. (1981), *Kategorien der Policy-Forschung*, Speyrer Arbeitshefte, 37.
- Karl, T.L. (1990), «Dilemmas of Democratization in Latin America», *Comparative Politics*, Vol. 23, Oct., pp. 1-22.
- Keatly, R. (1990), «Moral Medicine» *The Wall Street Journal/Europe*, March 5, 1,4.
- Keresztes, P. (1990), «Hungary Must 'Scrap and Build'», *The Wall Street Journal/Europe*, February 13.
- Kloten, N. (1989), «Zur Transformation von Wirtschaftsordnungen (The Transformation of Economic Systems).» *ORDO* 40, 99-127.
- Kloten, N. (1990), «Monetäre Steuerung in den transitorischen Stadien eines Wirtschaftssystems: Der Fall China. (Monetary Control in the Transitoric States of a Economic System.)» Erschienen in H.-J. Wagener *Monetäre Steuerung und ihre Probleme in unterschiedlichen Wirtschaftssystemen*, 191, Verein für Socialpolitik, Gesellschaft für Wirtschafts- und Sozialwissenschaften, 157-183.
- Kloten, N. (1990), *Die DDR-Wirtschaft im Systemwandel...* Stuttgart: Schabische Gesellschaft.
- Kloten, N. (1991), «Transformation einer zentralverwalteten Wirtschaft in eine Marktwirtschaft (The Transformation of a Centrally Planned to a Market Economic)», *Deutsche Bundesbank - Auszüge aus Presseartikeln* 1, January 4, 1-8.
- Kloten, N. (1991), «Die Transformation von Wirtschaftsordnungen. Theoretische, phänotypische und politische Aspekte», Walter Eucken Institut, Vorträge und Aufsätze Nr 132, Tübingen: Mohr Verlag.
- Kornai, J. (1990), *The Road to a Free Economy: Shifting from a socialist system again: the Example of Hungary*, New York: WW Norton.
- Koropecykj, I.S. (1991), Book Review on N.W. Balabkins (1988) *Not by Theory Alone: The Economics of Gustav von Schmoller and Its Legacy to America*, Berlin: Duncker & Humblot in *Eastern Economic Journal* 1, 140-142.
- Lowi, T.J. (1964), «American Business, Public Policy, Case Studies and Political Theory», in *World Politics*, Vol. 16, pp. 677-715.

- Lowi, T.J. (1972), «Four Systems of Policy, Politics and Choice, in *Public Administration Review*, pp. 298-310.
- Lowi, T.J. (1978), «Public Policy and Bureaucracy in the United States and in France», in Ashford D. (ed.), *Comparing Public Policies*, Amsterdam, (Sage Yearbook of Politics and Public Policy).
- Maier, H. (1991), «Integrieren statt zerstören. Für eine gemischtwirtschaftliche Strategie in den neuen Bundesländern», *Das Parlament*, Beilage: Aus Politik und Zeitgeschichte 29, 3-12.
- Mayer, J. (1990), «Focus of Outrage», *The Wall Street Journal/Europe*, March 13, 1,2.
- Mazmanian, D.A. and Sabatier P.A. (1981), *Effective Policy Implementation*, Lexington, Massachusettes.
- May, J.V. and Wildavsky, A. (1978), *The Policy Cycle*, London.
- Mayntz, R. (ed.) (1980), *Implementation politischer Programme*, Königstein, Taunus.
- Mayntz, R. (1983), *Implementation politischer Programme II*, Opladen.
- Mayntz, R. (1977), «Die Implementation politischer Programme: Theoretische Überlegungen zu einem neuen Forschungsgebiet», in *Die Verwaltung*, No. 1, pp. 51-78.
- Munchau, W. (1990), «New Germany faces innovation», *The Times*, October 4.
- Nikonov, V.A. and P.C. Ordeshook (1992), «Russia's Transition to Democracy: Essays 1-10», Working Paper 826 of the California Institute of Technology, Pasadena.
- Nikonov, V.A. and P.C. Ordeshook (1993), «Russia's Transition to Democracy: Essays 11-18», Working Paper 837 of the California Institute of Technology, Pasadena.
- N.n. (1990), «Bonn is urged to sell Assets to Finance Union», *International Herald Tribune*, October 4.
- N.n. (1990), «Bonn soll Staatsvermögen für die Einheit versilbern», *Frankfurter Neue Presse*, October 4.
- N.n. (1990), «Breuer: Bundesbank muß restriktiv bleiben», *VWD - Finanz und Wirtschaftsspiegel*, November 1.
- N.n. (1990), «Bundesbank: Aufwärts nach der Einheit», *Stuttgarter Nachrichten*, October 4.
- N.n. (1990), «Eine neue Privatisierungs-Runde: Bonn stellt die Weichen», *Czerwensky intern*, October 28-29.
- N.n. (1990), «Geld für Deutschland», *Frankfurter Allgemeine Zeitung*, October 4.
- N.n. (1990), «Größe Defizite bei Konvergenz», *Handelsblatt*, October 3.
- N.n. (1990), «Kanzler Kohls Kunst», *Frankfurter Allgemeine Zeitung*, May 21, 2.
- N.n. (1990), «40 Milliarden DM können eingespart werden», *Frankfurter Allgemeine Zeitung*, October 25.

- N.n. (1990), «Monetare Konfliktfelder der Weltwirtschaft», *Neue Züricher Zeitung*, October 10.
- N.n. (1990), «Schlesinger: Bundesbank fährt auf Sicht», *VWD - Finanz und Wirtschaftsspiegel*, October 3.
- N.n. (1990), «Schlesinger: Privatisierung für die Einheit», *Frankfurter Allgemeine Zeitung*, October 4.
- N.n. (1990), «Schlesinger: Privatisierungsforderung als Denkanstoß», *Frankfurter Allgemeine Zeitung*, October 18.
- N.n. (1990), «Slaughter the sacred cows», *The Economist*, November 24.
- N.n. (1990), «Staat soll Einheit durch Privatisierung bezahlen», *Frankfurter Rundschau*, October 4.
- N.n. (1990), «Ungewöhnliche Antworten», *Börsen-Zeitung Frankfurt*, October 5.
- N.n. (1990), «Zentralbank-Chefs diskutieren die Kontrolle von Allfinanz-Konzernen», *Die Welt*, October 5.
- N.n. (1991), «Der Bundesminister für Wirtschaft - Probleme der Privatisierung in den neuen Bundesländern», *Bundesanzeiger*, March 16, 1873-1877.
- N.n. (1991), «Survey - Business in Eastern Europe», *The Economist*, September 21, 3-30.
- N.n. (1994), «The Role of the Central Bank in Creating Viable Commercial Banks: Clearing Bank Balance Sheets & Other Issues», Meeting Report, *Institute for East-West Security Studies*, Economics 3.
- O'Donnell, G. and Schnitter, P. (1986), *Tentative Conclusions about Uncertain Transition*, Baltimore, John Hopkins Press
- O'Donnell, G., Schnatler, P. and Whitehead, L. (eds.) (1986), *Transition from Authoritarian Rule*, Baltimore, John Hopkins Press, 1986.
- Offe, C. (1975), *Berufsbildungsreform, eine Fallstudie über Reformpolitik*, Frankfurt.
- Rakoff, S.H. and Schaefer, G.F. (1970), «Politics, Policy and Political Science, Theoretical Alternatives», *Politics and Society*, Vol. 1, pp. 51-77.
- Ribhegge, H. (1992), «Der Sozialpakt für den Aufschwung von G. und H.-W. Sinn», *Wirtschaftsdienst* 3, 151-158.
- Rose-Ackerman, S. (1994), «Environmental Policy and Federal Structure: a Comparison of the United States and Germany», *Vanderbilt Law Review* 47(5), 1587-1622.
- Roth, T. (1990), «On your Mark», *The Wall Street Journal/Europe*, March 16-17; 1,2.
- Sabatier, P. (1975), «Social Movements and Regulatory Agencies: toward a more Adequate - and less Pessimistic - Theory of 'Clientele Capture'», in *Policy Sciences*, Vol. 6, pp. 301-327.
- Sabatier, P. and Mazmanian D. (1980), «The Implementation of Public

- Policy - A Framework of Analysis», in *Policy Studies*, pp. 538-560.
- Salisbury, R. (1968), «The Analysis of Public Policy», in Ranney A. (ed.), *Political Science and Public Policy*, Chicago, pp. 151-157.
- Salisbury, R. and Heinz J. (1970), «A Theory of Policy Analysis and Some Preliminary Applications», in Sharkansky, I. (ed.), *Policy Analysis in Political Science*, Chicago, pp. 39-60.
- Scharpf, F.W. (1976), «Theorie der Politikverfechtung», in Scharpf, Reissert, Schnabel, *Politikverflechtung: Theorie und Empirie des Kooperativen Föderalismus in der Bundesrepublik*, Kronberg.
- Schmidt, K. (1991), «Die Finanzierung des Einigungsprozesses in Deutschland», *Deutsche Bundesbank - Auszüge aus Presseartikeln* 67, September 11, 8-10.
- Schlesinger, H. (1990), «Die Wahrungspolitische Weichenstellungen in Deutschland und Europa», *Deutsche Bundesbank - Auszüge aus Presseartikeln* 78, October 5, 3-8.
- Schlesinger, H. (1991), «Wirtschafts- und Wahrungspolitik im Zeichen der Deutschen Einheit», *Deutsche Bundesbank - Auszüge aus Presseartikeln* 56, July 29, 1-5.
- Sharkansky, I. (1968), *Spending in the American States*, Chicago Rand McNally.
- Sharkansky, I. (1969), *The Politics of Taxing and Spending*, Indianapolis: Bobbs-Merrill.
- Sharkansky, I. (1970), *Regionalism in American Politics*, Indianapolis: Bobbs-Merrill.
- Sharkansky, I. (1970), *Policy Analysis in Political Science*, Chicago: Markham.
- Sharpe, L.J. and Newton, K. (1984). *Does Politics Matter? The Determinance of Public Policy*, Oxford: The Clarendon Press.
- Steinherr, A. (1991), «Essential Ingredients for Reforms in Eastern Europe», *Economic Journal on Eastern Europe and the Soviet Union* 3, 3-27.
- Thielbeer, S. (1990), «Ein Modell, das keines mehr ist», *Frankfurter Allgemeine Zeitung*, February 27.
- UNCTAD (1994), *Privatization in the Transition Process: recent Experiences in Eastern Europe*, United Nations Publication UNCTAD/GD/7.
- Veen, T. van (1990), «De eenwording van Duitsland», *Ikonomix* 5(5), 4-5.
- Wagner, H. (1991), «Einige Theorien des Systemwandels im Vergleich - und Ihre Anwendbarkeit für die Erklärung des gegenwärtigen Reformprozesses in Osteuropa», in: Jürgen Backhaus (ed.), *Systemwandel und Reform in östlichen Wirtschaften*, Marburg: Metropolis, pp. 17-40.
- Watrin, C. (1990), «Prämissen gesamtdeutscher Wirtschaftspolitik. (Premi-



- ses of The National German Economic Policy)», *Bund Katholischer Unternehmer e. V. Ausgewählte, Vorträge* 12.
- Wanniski, J. (1990), «The Germany's Rate a One-to-One Exchange Program», *The Wall Street Journal/Europe*, April 11, 6.
- Windhoff-Heritier, A. (1987), *Policy Analyse, eine Einführung*, Frankfurt.
- Wiseman, J. and J. Casas Pardo (1991), «The Political Economy of the Spanish 1978 Constitution Making Process», paper presented at the Annual Meeting of the European Public Choice Society, Beaune, April 3-6.
- Yunker, J.A. (1992), *Socialism Revised and Modernized*, Westport: Greenwood Publishing Group.
- ZDF - Mittagmagazin (1990), Interview mit Prof. Dr. H. Schlesinger, October 15.





## The military revolution at sea<sup>1</sup>

*Nicolas Rodger\**

### *Abstract*

*The article examines the link between the nature of societies, the nature of governments, and the nature of their military activities. Three great military and economic powers of the seventeenth and eighteenth centuries are compared, Britain, Sweden and the Netherlands. It is concluded that seapower was most successful in countries with flexible and open social and political systems.*

Historian and political scientists have long been interested in the rise of the modern state, and have very frequently connected it with what Michael Roberts called the 'Military Revolution'. At its crudest, the idea he popularised is that at a date usually located in or around the sixteenth century, armies became very much larger and more costly, subjecting the states which sought to raise them to severe strains. Many early modern states failed to meet this challenge, were conquered, absorbed or marginalised. A few succeeded in generating the permanent tax-raising powers needed to sustain modern armies, and so become the great military powers of modern Europe. These powers –France, Prussia, Sweden, sometimes Russia, Austria and even Spain are usually cited– are often described as having followed a particular route to modernity, sweeping away the archaic medieval representative institutions which had hampered the power of princes, and becoming autocratic, centralised and bureaucratic po-

---

\* Professor of History, University of Exeter.

wers in which the power of the state had the full play necessary to meet the military and governmental challenges of the modern world.

There seems to be a number of difficulties inherent in this approach. By focusing attention on the rise of the great powers, it implicitly assumes that only power and success are of interest to history. Political scientists in particular, seeking to identify the two (or, in the more sophisticated versions, three) variables which explain the rise of the great powers, have largely ignored the experience of states and nations which did not succeed, or succeeded in different ways.<sup>2</sup> Moreover the concept of the 'military revolution' explains the fate of nations by reference to the form and functions of government. It assumes that the state shaped society, rather than society the state. Finally, historians have with few exceptions considered only armies, not navies.<sup>3</sup> This is beginning to change, but it is still possible to read eminent historians comparing Britain and Prussia as military powers who are apparently quite unaware that the British navy had any historical importance, or indeed had ever existed.<sup>4</sup>

Broadening the perspective to consider small and transient states as well as great powers, naval as well as military influences on them, and society as much as government, draws our attention to states which seem to have been formed in very different ways. Genoa, for example no longer survives as an independent state, let alone as a mother of states, and yet medieval Genoese seapower gave rise to several naval republics. Some of them were not very far from here, on Aegean islands;<sup>5</sup> and one of them, Monaco, still survives as a mute witness to the Genoese seapower of the thirteenth and fourteenth centuries.<sup>6</sup>

Considering states which never became great powers teaches us, among other things, that modernity and bureaucracy in central government are not necessarily sufficient for success. In 1588, for example, one of the best-known of European states radically reformed its system of government, sweeping away the untidy mixture of offices and institutions inherited from the Middle Ages, and replacing them with a coherent and logical system of fifteen boards or departments which divided the business of government on functional lines, each department headed by a minister responsible to his sovereign for its affairs. It was then a unique approach to government, and it has since become universal, which makes it curious that so influential a development has been so little studied. The state in question is of course the Papal State, and it is Sixtus V's great constitution *Immensa aeterni Dei* which ushers in the modern era of government.<sup>7</sup> If the organisation of the state was what

was required to make a great power, the Pope's divisions should have dominated Europe. Reality, as we know, was different. The Papal State, admirably equipped for greatness at the level of central government, nevertheless lacked many of the essential requirements of a modern state.<sup>8</sup>

Just as the Papal State was being refounded, another European state was struggling into existence, and organizing its nascent naval forces, with very few of the blessings of good government provided by Sixtus V. The Dutch Republic in over two centuries of existence never managed to agree to permanent constitution to replace the temporary political arrangements with which it had begun. Its navy, or rather navies, consisted of a large number of more or less independent organizations.<sup>9</sup> Five provincial admiralties (notionally federal institutions, but in practice dominated by local interests) each maintained from their own revenues a fleet, an establishment of officers, and one or more naval yards. The two great joint stock companies, the East and West India Companies, each possessed substantial numbers of men-of-war. A number of individual seaports provided municipal navies (the 'directieschepen') to protect their own shipping. Private syndicates, especially in Zealand, commissioned squadrons of privateers to prey on enemy shipping. None of these were directly controlled by the central government – indeed it is not obvious that the 'Generaliteit' of the Dutch Republic can be called a central government in the ordinary acceptance of the phrase, when it had no ministers, or ministries, and hardly any permanent national institutions. Sovereign authority resided in the States-General, when it was in session, but the assembly itself consisted of delegates (not representatives) who could not act without a mandate from their principals, the provincial estates, which in turn were obliged to consult their electors before reaching a decision. No political scientist could possibly accept such a government as a plausible candidate for the military revolution, or even the modern world. Yet it was this system which created one of the foremost of the new armies of the sixteenth century, and by 1639 had made the United Provinces the leading European naval power. Not until 1654 did it acquire its first national fleet, paid for out of national revenues and unequivocally the property of the Republic – though even then the fleet formed but a part of the available naval forces, and its actual building and operation was perforce entrusted to the provincial admiralties, since the state possessed no other naval administration. Such a republic, where government was an abstract noun, where authority

was endlessly divided and disputed, seems to historians today and seemed to many observers then to be completely unfitted for survival in a competitive world. Yet here again, theory and practice were very different.

By the mid-seventeenth century, there was another great power in Europe, one which deployed a first-class army and navy in spite of lacking most of the obvious requirements of a great power. Sweden was a poor and marginal country with a small population and no significant shipping interest. So far from being dominated by a powerful monarch, Sweden experienced a twelve-year regency from 1632 to 1644, followed by the reign of a woman, Queen Christina. Certainly her father Gustaf Adolf had been a great military leader and tactical innovator, but it was not absolutist power which enabled him to mobilize Sweden's resources. On the contrary, the rise of the Vasa dynasty to national and international power depended on creating a representative institution, the Riksdag, which incorporated the aristocracy and gave them a strong interest in the state. This was not so much an imposition of royal power as a dispersal of it. The state also developed efficient systems of local taxation and military mobilisation which were built from below rather than imposed from above, and which rested on a large measure of co-operation and consensus. In the 1640s and 50s, at the same time as aristocratic and popular revolts were tearing apart France, Spain and England, the Swedish political system was smoothly coping with a royal minority, the rule and subsequent abdication of a queen, and the accession of a new dynasty. At the same time it made possible a high level of effective taxation, by which a poor country with a small population supported the army and navy of a first-class power.<sup>10</sup>

It has not entirely escaped historians' notice that the Dutch, the Swedes and finally the British rose to the status of great powers; the Dutch without much that could be described as a central government, the British without much that could be called an army, and none of them with large populations or despotic forms of government. Attempts have therefore been made to explain, or at least to explain away, these apparent exceptions from the rule that the demands of the military revolution created centralised, autocratic or bureaucratic military powers. The sea itself has always been a favourite explanation, allegedly making England 'invasion-proof', and allowing it the luxury of retarded development until at length it became wealthy enough to support both a navy and a first-class army – the latter implicitly constituting the admission ticket to the great

powers' club.<sup>11</sup> Unfortunately there are several difficulties with this theory. In reality the sea was never in the least impassable; certainly not by the standards of early modern roads. Until, and indeed for some time after, the English developed effective naval forces, the country was repeatedly invaded by sea. English governments have been overthrown by seaborne invasions at least nine times since the Norman Conquest: in 1139, 1153,<sup>12</sup> 1326, 1399, 1460, 1470, 1471, 1485 and 1688; to which should be added the 1332 invasion of Scotland, and at least seven other successful landings of major forces in England (in 1069, 1101, 1215, 1405, 1462, 1469 and 1487) which went on to campaign but did not overthrow the regime. These figures take no account of lesser raids and landings, or of seaborne assistance against England sent to Wales, Scotland or to English rebels; they ignore all expeditions which did not succeed in putting troops ashore, and they do not include landings of any kind in Ireland.<sup>13</sup> Countries like Spain which were protected by high mountains might genuinely claim to be insulated by geography from the military revolution, but England could not, and still less Sweden or the Netherlands. Nor is it in the least persuasive to argue that the English navy did not exist, or did not cost very much, or did not experience any significant technical developments, until the country was already a wealthy great power.<sup>14</sup> On the contrary, it is abundantly clear that a real 'naval revolution', both technical and administrative, took place in England in the mid-sixteenth century, which posed an enormous challenge for the essentially medieval structure of English government.<sup>15</sup> Though Queen Elizabeth's income was no higher in real terms than that of her medieval predecessors, in wartime she regularly spent one-third of it on her Navy,<sup>16</sup> and this high expenditure was no passing fancy but a sustained long-term national policy. It was established well before England had a large merchant fleet, or any significant oceanic trade, or any distant colonies. Certainly naval expenditure placed immense strains on the English political system, and contributed substantially to its collapse in the 1640s. In that sense England was indeed one of the casualties of the military revolution.<sup>17</sup> In the event, however, the country, and its constitution, and its navy, survived the crisis and emerged from it greatly strengthened, but even less autocratic than before. The Dutch had meanwhile become the world's leading naval power, surmounting in the process immense technical, administrative and financial obstacles, entirely without a national navy and almost without a national government. What England, the Netherlands and Sweden had in

common was a structure of government which required an unusually large degree of consent and co-operation by the governed, and permitted a much higher level of effective 'resource extraction', through the medium of taxation, loans and private participation in the national war effort.

We must be absolutely clear that maintaining even a small permanent navy was (and indeed still is) extremely complex, costly and demanding. The industrial, technical and managerial resources required to build and operate warships vastly exceeded in kind and quality anything needed by an early modern army. In the 1560s the manufacture of hand guns was still a cottage industry and the casting of heavy guns (invariably in bronze) a luxury business for the gratification of princes. Yet in England the Ordnance Board had already spent thirty years and considerable sums of money in developing a large private iron gun founding industry. In an age when the largest private enterprises employed fewer than a score of people, the English dockyards were already employing six or seven hundred, in a wide range of skilled trades, with a substantial managerial and administrative structure. Even the largest sixteenth-century armies did not call for anything like the skills, the capital investment and the long-term commitment of a small navy.<sup>18</sup> It is no coincidence that Sweden, in most respects a peasant economy on the margins of international trade, had by the mid-seventeenth century developed one of the most advanced iron, and in particular gunfounding, industries in Europe (though this was greatly the work of Dutch investment, and exported its guns to arm Dutch ships). It was navies, not armies, which first confronted the demands of the industrial age.

The needs of an early-modern army were essentially those of mobilizing manpower on a large scale. The first requirement was a large number of unemployed peasants to make soldiers, and a smaller number of unemployed noblemen to make officers. The actual raising, training and, often, the equipping of the troops was left to the regimental officers. The feeding and paying them was as far as possible laid on the populations of conquered or occupied territories, to such an extent that in the (admittedly extreme) case of Swedish armies during the Thirty Years' War, it was even possible to wage major campaigns at a notional profit.<sup>19</sup> The huge fortresses of the artillery age required mass manpower and a good deal of masonry and timber revetment, but in engineering terms they were hardly larger or more sophisticated than the hill-forts of the Iron Age.<sup>20</sup> Raising the great armies of the early-modern age certainly imposed



costs and strains on government, but they did not present a new challenge to society. They called for noblemen to do, and peasants to bear, what noblemen had always done and peasants had always borne. In social terms, a large seventeenth or eighteenth-century army was not very different from a small thirteenth or fourteenth-century one. These armies belonged naturally to societies which had not changed, and did not intend to change the conservative social order, societies in which the three estates kept to their God-given conditions, societies built around throne and altar.<sup>21</sup> 'The military hierarchy reproduced the fundamental social hierarchy, with all its privileges and inequalities.'<sup>22</sup> If this was a military revolution, its effects on society were the opposite of revolutionary. Government may have been changed, at least in some countries, by the development of modern bureaucracy, but society was ossified rather than revolutionised. The military revolution may have equipped absolutist monarchies to face the relatively simple challenge of mobilizing mass manpower, but it is not at all clear that it did anything to prepare them for the industrial world.

The great survivor, of course, was the great failure: Britain, the country which failed to rise to the challenge of the military revolution.<sup>23</sup> The question for the historian is how and why Britain survived, and there does not seem to be any agreed answer. It has been argued that British government or British bureaucracy were uniquely efficient,<sup>24</sup> and strikingly inefficient;<sup>25</sup> that the country was surprisingly militarized, and unusually free of military influence. It has been described as unique in combining the 'urban, capital-intensive' path to modernity with a strong central government.<sup>26</sup> For some scholars, England was different because it had a strong Navy, and a strong Navy was what made England different.<sup>27</sup> For others England was different because it had a strong Parliament, and a strong Parliament was what made England different.<sup>28</sup> Neither observation seems to have quite the explanatory force we need, and in any case weighty scholarly opinion argues that the English Parliament in the seventeenth century was just as weak as the English Crown, and weaker than its Continental analogues.<sup>29</sup>

It is not original to suggest that navies and constitutional government go together as naturally as armies and absolutism. Aristotle was quite clear on the point,<sup>30</sup> and other political scientists have followed him.<sup>31</sup> It is a matter of observation that army officers have traditionally been noblemen or gentlemen, while navies have tended to be run by middle-class professionals on whom the aristocratic

concept of honour sat somewhat awkwardly.<sup>32</sup> What has not, to my knowledge, been seriously examined is why this should be so. Is it coincidence, or is it in the nature of a navy to favour constitutional rather than autocratic government? If there is a connection, which is cause and which effect? How far can this line of enquiry explain the British case? My answer starts from the argument of Jan Glete that successful navies require the support of a coalition of 'interest groups', united in supporting a strong navy, and capable of translating that support into long-term political commitment.<sup>33</sup> This argument has force regardless of the type of government in question, and one of its obvious merits is that it tends to draw attention away from the form of central government towards the nature of society, or at least of the political classes of society. Glete is quite clear that in the early modern period at least, efficient navies could be organised by provincial government or private individuals, though in the long run only states could meet the huge cost of major fleets.<sup>34</sup> Indeed this argument could be taken much further back, for in the early Middle Ages England, the Scandinavian countries and some parts of the Celtic world organised very large fleets by the *leidang* or ship-muster system, which depended on societies possessing advanced technical capabilities in shipbuilding and shiphandling, and a well-developed sense of mutual obligation and common purpose, but hardly on central government at all.<sup>35</sup>

Glete's argument helps to explain one of the obvious paradoxes of navies and state-formation. The examples of Spain in the late sixteenth century, England in the mid-seventeenth, France in the late seventeenth, Germany in the late nineteenth, and Russia in more than one period, all show that autocratic, militarised states are perfectly capable of building large and efficient navies, often with astonishing speed. What they do not seem to be capable of is sustaining their creations. Spanish seapower enjoyed a brief period of strength in the 1590s followed by a steep decline. The English Republic (and the English army which dominated it) took barely ten years to create the most formidable navy in Europe, and then to collapse. Louis XIV's fleet rose to be the largest in the world in less than thirty years, and had largely disappeared within another thirty. The fleet that Tirpitz built on borrowed money ran out of credit in the budget crisis of 1912. All these cases can be well explained by the argument that the temporary influence of a dominant favourite or the capricious will of the All-Highest were no substitute for the solid support of entrenched interest groups.<sup>36</sup>

It may be, however, that the argument needs to be taken further, for there is clear indication that absolutist governments were not merely bad at assembling a broad political constituency to support seapower; they seem actually to have had structural difficulties in sustaining a navy and using it rationally, regardless of the degree of government support.<sup>37</sup> In the case of Spain the stresses of war, and particularly war at sea, corroded the fabric of the absolutist state. More and more of the functions of government, including the function of building and operating fleets, were ceded to private contractors and private interests because the state was incapable either of organising or paying for them. At the same time the burdens imposed by the state largely contributed to ruining the country, impoverishing agriculture, destroying industry and reducing the largest merchant fleet and shipbuilding industry in the world to a miserable condition from which they have scarcely since recovered.<sup>38</sup> Perhaps the case of Spain is exceptional, but in the long run Oliver Cromwell had little more success than Olivares. Born of the insecurity and vulnerability of a minority regime, floated on extraordinary revenues which could not be sustained,<sup>39</sup> the English fleets of the 1650s lacked any the conditions for long-term survival. In the event the Republic collapsed before its navy could, leaving the restored monarchy to continue the search for a long-term foundation of seapower. Charles I with his Ship-Money fleets, Charles II with the Third Dutch War, both tried to use their navies as a lever for establishing something like absolutist government, and both found the instrument working to overturn their own ambitions. Even France under Louis XIV, which enjoyed unequalled revenues and apparently complete political stability, seems to have needed more than political commitment to endow its impressive fleet with lasting vigour. The latest study bluntly concludes that the navy's status as the personal creation of Colbert fatally weakened it from the start, making it an instrument of court politics rather than national policy. What was worse, its effectiveness was compromised precisely by its bureaucratic character as an accountant's navy, dedicated to efficiency rather than effectiveness: a sign of modernity to gladden the heart of Max Weber, no doubt, but an obstacle to victory, which so often depends on the deliberate sacrifice of efficiency.<sup>40</sup> Certainly the eighteenth-century French navy, even in the periods in which it enjoyed political backing and adequate finance, always lacked many of the essential elements of seapower, both human and material.<sup>41</sup> Always it remained a creature of the state, every element of it de-

pendent on the support of the state. If foundries ran out of orders for anchors or guns, they closed, for they had no private customers. If the dockyards could not build enough ships of the line, there were no private yards to turn to. A century's effort to purge the officer corps of bourgeois, Protestants and other social undesirables turned it into the perfect expression of the absolutist regime, but a very imperfect instrument for winning wars.<sup>42</sup>

In all of these cases, moreover, the power of the absolutist or military state to impose taxation to pay for its naval ambitions was a key to success in the short term, and disaster in the long term. Absolute monarchies could destroy themselves, either by laying intolerable burdens on their economies, as in the case of Spain; or by incurring intolerable debts, as in the case of seventeenth-century England and eighteenth-century France.<sup>43</sup> Then as now, it seemed very much quicker and more efficient for an enlightened ruler to impose his will than for a pluralist government to assemble a coalition in support of a policy. Only a broad coalition, however, could ensure that the burden of war was distributed in the most equitable and least damaging fashion. Only a system capable of arousing and expressing popular support could make possible the high levels of effective taxation necessary to support industrial warfare.<sup>44</sup>

In Britain the Navy was always shaped by the nature of society as much as the will of the state. The English navy from its Tudor origins depended on a close alliance between the Crown and private interests. The naval war united nobility, gentry, merchants and seamen in 'the directest expression of the nation-in-arms'.<sup>45</sup> The Navy Royal was part of this coalition, and not exclusively under the control of the Crown. This has been described as a consequence of the weakness of the state, which in a way it was; and as a kind of decay or corruption, 'a disintegration of power, resulting from the conjunction of an antiquated system of government and the acquisitive drive of vigorous private interests';<sup>46</sup> which is true only from a quasi-Marxist perspective which sees the state as the only true expression of society and the only legitimate instrument of military force. Even when the English (later British) state became more efficient and wealthier, in the late seventeenth and early eighteenth centuries, the Navy and the state continued to rely heavily on private interests to perform functions which in France or Spain were the monopoly of the state. Private industry cast guns and manufactured powder, provided all sorts of stores and equipment and built many of the ships. Senior officers were almost entirely responsible for selecting

and training their own officers and men; not until well into the nineteenth century did the Admiralty gain control of the entry of officers into the Navy.<sup>47</sup> It is true that there were important elements of naval administration which were more centralised in Britain than in France, notably victualling and central finance, but the French *trésoriers-généraux* and *munitionnaires* were not private contractors operating in a free market so much as privileged monopolists exploiting a private relationship with the centres of power.<sup>48</sup> They represented key points of weakness in French naval administration, where the British Navy's extensive networks of private suppliers and contractors gave it resilience and formed so many roots anchoring the Navy in the soil of civil society. Above all, 'Dutch finance' in both its Amsterdam and London versions, invoked private skills and public markets to spread the financial burden and risk of war across the whole breadth of the propertied classes at home and abroad. The funded debt represented the political nation's solid commitment to the wars which were a real expression of national will, not the private ambitions of kings or ministers. Two important recent studies have now shown in detail how Britain rose as a 'fiscal-military' state, combining a remarkably high and sustained level of taxation with an open and representative system of government. Though the essential technical advances in taxation and borrowing were made by the English republic in the 1650s, they were continued by the restored Parliament of the Restoration, and developed in the late seventeenth and early eighteenth centuries.<sup>49</sup>

Comparing Britain, Sweden and the Netherlands with the other great powers of the seventeenth and eighteenth centuries suggests that open and pluralist societies were not only better at assembling political support for navies, but better at the organization of seapower. To integrate the wide range of human, industrial, technical, commercial and managerial resources required to build, maintain and fight a seagoing fleet was simply beyond the unaided competence of any state in the early modern period, and perhaps it still is. Nations in which public policy was based on a broad consensus of interests, in which numerous private businesses serviced and influenced government, in which land and trade overlapped, were much better equipped to sustain a navy.<sup>50</sup> Effective seapower was the product of society as well as the state – if not instead of the state, in the case of the Dutch Republic. Only flexible and integrated societies could surmount the very considerable difficulties involved. Money was certainly necessary, but money alone did not

suffice. The events of 1588 amply demonstrate that even ten times the financial resources of England were no substitute for the efficient administration and extensive networks of support from private industry and trade which Spain lacked.<sup>51</sup>

This is a statement about the nature of society as much as about the form of government. Absolutist government had difficulty sustaining effective seapower because it had difficulty mobilizing the broad coalition of interests needed to support so complex an enterprise. Indeed it did not wish to mobilize many elements in society, either because they were actively undesirable (Moriscos in Spain, Huguenots in France, Jews in many countries), or because the absolutist system did not allow them freedom of action, or because it did not wish to involve them in the business of the state. Merchants, industrialists, tradesmen, skilled craftsmen; all had a place in absolute monarchies, but it tended to be a subordinate or marginal place. Indeed it is striking how often in absolutist societies the place of trade and industry was literally marginal; geographically remote from inland capitals, on the frontiers or seacoasts of the state.

All this suggests a link between the nature of societies, the nature of governments, and the nature of their military activities. Middle-class participation in public life, professional skills, commerce, industry and private finance directly favoured and were favoured by navies. Seapower was most successful in countries with flexible and open social and political systems.<sup>52</sup> They were the same which favoured trade and industry, and for the same reason, for a navy was the supreme industrial activity. The armed forces of early modern states were the blueprint of their modern societies: a complex, integrated, industrial world for the naval powers; a rigid, archaic world of great landed estates for the military powers.<sup>53</sup> It has been argued that a Europe of warring states eventually came to dominate a world much of which was occupied by vast territorial empires, precisely because the ceaseless competition of European warfare bred forces and weapons which were much more capable than the products of closed, monopolistic societies – in short, that the free market was more efficient than the plan.<sup>54</sup> If this was true on an international level, it should have been equally true within nations. If free competition is more economically efficient than government direction, then those forms of military activity which best harmonise with free markets are likely to be more successful in themselves, and best promote (or least hamper) the growth of flexible and competitive economies. In a world which was steadily moving away from land and population as the sole sources of wealth and towards commerce

and industry, naval powers and open societies were much better placed to adapt than military powers and autocratic monarchies. It was not simply that sea-power gave access to overseas wealth, important though that was, for absolutist states could and did elect to become seapowers. They, however, were conspicuously less successful than other states, not richer than they and in some cases much poorer, but better equipped to mobilize their whole resources, either for war or for peace. The critical difference was the nature of society, and government as an expression of society, rather than constitutional forms taken in isolation. Open societies were best at naval warfare for the same reason that they were later best at meeting other challenges of the modern world, because a navy was an image of the modern world in miniature. 'Warfare on the British model was a triumph for an enterprising and acquisitive society, not an authoritarian one.'<sup>55</sup> Britain did not simply survive centuries of warfare relatively unscathed because of geographical and historical accident,<sup>56</sup> to profit from the industrial revolution because there were no competitors left undevastated by war. Naval warfare was Britain's apprenticeship for commercial and industrial supremacy.

## Notes

1. In its original form this paper was first given at the Canadian Nautical Research Society's International Conference on Maritime History, Calgary, 1998. I am indebted to Professor Jan Glete for suggestions and references which have greatly improved it.

2. See for example: Richard Bean, 'War and the Birth of the Nation State', *Journal of Economic History* XXXIII (1973) pp. 203-221; Samuel E. Finer, 'State and Nation-Building in Europe: The Role of the Military', in *The Formation of National States in Western Europe* ed. Charles Tilly (Princeton, 1975) pp. 84-163; Ronald W. Batchelder & Herman Freudenberger, 'On the Rational Origins of the Modern Centralized State', *Explorations in Economic History* XX (1983) pp.1-13; Charles Tilly, 'War Making and State Making as Organized Crime', in *Bringing the State Back In* ed. Peter B. Evans, Dietrich Rueschmeyer & Theda Skocpol (Cambridge, 1985) pp. 169-191. The most recent contributions are: *Power Elites and State Building* ed. Wolfgang Reinhard (Oxford, 1996); Thomas Ertman, *Birth of the Leviathan: Building States and Regimes in Medieval and Early Modern Europe* (Cambridge, 1997); and *idem*, 'The Sinews of Power and European State-Building Theory', in *An Imperial State at War: Britain from 1689 to 1815* ed. Laurence Stone (London, 1994), pp. 33-51.

3. This applies to most of the very numerous books on the subject. Honourable exceptions are: Geoffrey Parker, *The Military Revolution: Military Innovation and the Rise of the West, 1500-1800* (Cambridge 1988); and Jeremy Black, *A Military Revolution? Military Change and European Society 1550-1800* (London, 1991). Jaime Vicens Vives, 'Estructura administrativa estatal en los siglos XVI y XVII', *Rapports IV, XIe Congrès International des Sciences Historiques*, (Stockholm, 1960) pp. 1-24 (also in *Obra Dispersa* ed. M. Batllori & E. Giralt (Barcelona, 1967, 2 vols) II, 359-377; and translated by Frances M. Lopez-Morillas as 'The Administrative Structure of the State in the Sixteenth and Seventeenth Centuries', in *Government in Reformation Europe, 1520-1560* ed. Henry J. Cohn (London, 1971) pp. 58-87), was perhaps the first to mention the naval contribution to state formation. Jan Glete, *Navies and Nations: Warships, Navies and State Building in Europe and America, 1500-1860* (Stockholm, 1993, 2 vols) is an outstanding study which eschews the 'military revolution' as such; Professor Glete has now followed it with *War and the State in Early Modern Europe: Spain, the Dutch Republic and Sweden as fiscal-military states, 1500-1660* (London, 2002). Brian M. Downing, *The Military Revolution and Political Change: Origins of Democracy and Autocracy in Early Modern Europe* (Princeton, 1992), pp. 72-73, 165 & 224, is a rare political scientist who is uneasily aware that the existence of navies somewhat undermines his argument. *War and Competition between States* ed. Philippe Contamine (Oxford, 2000) has a chapter by Jaap R. Bruijn, 'States and their Navies from the Late Sixteenth to the End of the Eighteenth Centuries', pp. 69-98.

4. *Rethinking Leviathan: The Eighteenth-Century State in Britain and Germany* ed. John Brewer & Eckhart Hellmuth (Oxford, 1999).

5. William Miller, 'The Gattilusi of Lesbos, 1355-1462', *Byzantische Zeitschrift XXII* (1913), pp. 406-447. Michel Balard, *La Romanie Genoise* (Genoa, 1978, 2 vols); *idem*, 'Les Genoës en mer Egée', *Mediterranean Historical Review IV* (1989), pp. 158-174.

6. Gustave Saige, *Monaco, ses origines et son histoire* (Monaco, 1897), pp. 42-56. Léon-Honoré Labande, *Histoire de la Principauté de Monaco* (Monaco, 1934), pp. 27-44. Michel Balard, 'Course et piraterie à Genes à la fin du Moyen Age', in *El comerç alternatiu: corsarisme i contraban (SS.XV-XVIII)* ed. Gonçal Llorens Nadal (Palma de Mallorca, 1990) pp. 29-40, at p. 33.

7. Pio Pecchiai, *Roma nel Cinquecento* (Bologna, 1949), pp. 173-177. Paolo Prodi, *The Papal Prince. One Body and Two Souls: The Papal Monarchy in Early Modern Europe* trans. Susan Haskins (Cambridge, 1987), p. 87. Niccolò Del Re, *La Curia Romana* (Rome, 2nd edn 1952), pp. 19-20 & 175-176. Jean Delumeau, 'Le progrès de la centralisation dans l'État pontifical au XVIe siècle', *Revue Historique* 226 (1961) 399-410, at p. 403. The text is printed in *Bullarium Romanum Novissimum* ed. L. Cherubini et al. (Rome, 3rd edn 1638, 4 vols) II, 464-468. Curiously enough these authorities do not agree on the actual date of the bull. Most modern authorities (e.g. Prodi,



p.87; Del Re, p.87) plump for 22 Jan 1588; Alberto Guglielmotti, *Storia della Marina Pontificia* (Rome, 1886-93, 10 vols) VII, 22, prefers the same day in 1587, Pecchiai (pp.176 & 207) offers both, while the actual text in Cherubini's (indifferent) edition is dated 1 Feb 1587.

8. Cf Delumeau, 'l'État pontifical', p. 410: 'Ce qui a manqué à l'État pontifical de la Contre-Réforme, ce ne sont pas tant les structures administratives que la santé sur le plan économique et social.'

9. What follows is a substantial simplification. Good introductions to Dutch naval administration are J.R. Bruijn, *Varend Verleden: De Nederlandse Oorlogsvloot in de 17e en 18e eeuw* (Amsterdam, 1998), and the appropriate sections of the *Maritieme Geschiedenis der Nederlanden* ed. G. Asaert et al. (Bussum, 1976-78, 4 vols). Glete, *War and the State*, pp. 140-173, considers Dutch government and war more generally.

10. Jan Glete, 'La construcción de un Impero con recursos limitados: Suecia y el desarrollo de las organizaciones militares', in *España y Suecia en la época del Barroco (1600-1660)* ed. E. Martínez Ruiz & M. de Pazzis Pi Corrales (Madrid, 1998), pp. 307-339. Glete, 'Absolutism or Dynamic Leadership? The rise of large armed forces and the problem of political interest aggregation from a mid-17th-century perspective', in *Politics and Culture in the Age of Christina* ed. Marie-Louise Roden (Stockholm, 1997), pp. 23-28. I am indebted to Professor Glete for copies of these papers, now conveniently summarised in his *War and the State*, pp. 174-212.

11. Perry Anderson, *Lineages of the Absolutist State* (London, 1974), p. 123. John Brewer, *The Sinews of Power: War, Money and the English State, 1688-1783* (London, 1989), pp. 12-13. *The Medieval Military Revolution: State, Society and Military Change in Medieval and Early Modern Europe* ed. Andrew Ayton & J.L. Price (London, 1995), p. 6. Downing, *Military Revolution*, pp. 179-183.

12. Perhaps 'seriously undermined' would describe these two occasions better than 'overthrown'.

13. N.A.M. Rodger, *The Safeguard of the Sea. A Naval History of Britain*, Vol. I, 660-1649 (London, 1997), p. 429.

14. Michael Mann, *The Sources of Social Power. I: A History of Power from the Beginning to A.D. 1760* (Cambridge, 1986), p. 457. Cf the comments of Ertman, *Birth of the Leviathan*, pp. 14-19.

15. Ronald Pollitt, 'Bureaucracy and the Armada: the Administrator's Battle', *Mariner's Mirror* LX (1974) pp. 119-132.

16. P.K. O'Brien & P.A. Hunt, 'The Rise of a Fiscal State in Britain, 1485-1815', *Historical Research* LXVI (1993) pp. 129-176. Geoffrey Parker, 'The Dreadnought Revolution of the Sixteenth Century', *Mariner's Mirror* LXXXII (1996) pp. 269-300.

17. Kenneth R. Andrews, *Ships, Money and Politics: Seafaring and Naval Enterprise in the Reign of Charles I* (Cambridge, 1991), p. 3. Conrad Russell, *Parliaments and English politics, 1621-1629* (Oxford, 1979) pp. 64,

418 & 431. Simon Adams, 'Spain or the Netherlands? The Dilemmas of Early Stuart Foreign Policy', in *Before the English Civil War: Essays on Early Stuart Politics and Government* ed. Howard Tomlinson (London, 1983) pp. 79-101, at p. 80.

18. Rodger, *Safeguard of the Sea*, pp. 228-237.

19. Downing, *Military Revolution*, pp. 193-202. Parker, *Military Revolution*, pp. 66-67. But Professor Glete is clear that the Swedish state (as distinct from generals and military entrepreneurs) never made a true profit.

20. I have failed to find directly comparable figures for the quantities of earthwork and masonry required, but there are some suggestive calculations in *The Experimental Earthwork on Overton Down, Wiltshire, 1660* ed. P.A. Jewell (British Association for the Advancement of Science, London, 1963), pp. 51-58, and John Coles, *Archaeology by Experiment* (London, 1973), pp. 73-74.

21. This is true regardless of whether you believe the military revolution was the cause of absolutism, or (like Black, *Military Revolution*) the consequence.

22. Christopher Storrs & H.M. Scott, 'The Military Revolution and the European Nobility, c. 1600-1800', *War in History* III (1996) pp. 1-41, at p. 34.

23. Jean Meyer, *Le poids de l'état* (Paris, 1983), pp. 128-129.

24. Brewer, *Sinews of Power*, *passim*.

25. Wallace T. MacCaffrey, Elizabeth I, *War and Politics 1588-1603* (Princeton, 1992), pp. 23-41. Lawrence Stone, introduction to *An Imperial State at War: Britain from 1689 to 1815* (London, 1994), pp. 14-17.

26. Charles Tilly, Coercion, *Capital and European States, AD 990-1990* (Oxford, 1990), pp. 56-61.

27. Aristide R. Zolberg, 'Strategic Interactions and the Formation of Modern States: France and England', in *The State in Global Perspective* ed. Ali Kazancigil (Paris & Aldershot, 1986) pp. 72-106, at pp. 94-95.

28. This seems to sum up the argument of Ertman, *Birth of the Leviathan*.

29. C.S.R. Russell, 'Monarchies, Wars and Estates in England, France and Spain, c. 1580-c. 1640', *Legislative Studies Quarterly* VII (1982) pp. 205-220.

30. *Politics* VI. 6 "2.

31. Mann, *Sources of Social Power*, I, 478-481. Anderson, *Lineages of the Absolutist State*, p. 134.

32. N.A.M. Rodger, 'Honour and Duty at Sea, 1660-1815', *Historical Research* LXXV (2002) pp. 425-447.

33. Glete, *Navies and Nations* II, 477-489.

34. Glete, *Navies and Nations* I, 159-161.

35. Hans Kuhn, *Das Altnordische Seekriegswesen* ed. Sigrid Engeler & Dietrich Hofmann (Heidelberg, 1991), pp.54-87. Nicholas Hooper, 'Some Observations on the Navy in Late Anglo-Saxon England', in *Studies in Me-*

dieval History presented to R. Allen Brown ed Christopher Harper-Bill et al. (Woodbridge, 1989), pp. 203-213. *Kulturhistorisk Leksikon for Nordisk Middelalder* (Copenhagen &c, 1956-78, 22 vols) X, 433-446, s.v. Leidang. Lucien Musset, 'Problemes militaires du monde scandinave (VIIe-XIIe siecles)', in *Ordinamenti Militari in Occidente nell'alto Medioevo* (Spoleto, 1968) pp. 229-291, at pp. 279-284. Else Roesdahl, *Viking Age Denmark* (London, 1982), p. 157. Klavs Randsborg, *The Viking Age in Denmark: The Formation of a State* (London, 1980), pp. 32-43. Niels Lund, 'The Armies of Swein Forkbeard and Cnut: *leding or lið?*', *Anglo-Saxon England XV* (1986) pp. 105-118. John Bannerman, *Studies in the History of Dalriada* (Edinburgh, 1974), pp. 140-141. Hugh Marwick, 'Naval Defence in Norse Scotland', *Scottish Historical Review XXVIII* (1949) pp. 1-11.

36. Martine Acerra & Andre Zysberg, *L'essor des marines de guerres europeennes (vers 1680 - vers 1790)* (Paris, 1997), p. 53. Michel Vergé-Franceschi, *La marine française au XVIIIe siècle: Guerres, Administration, Exploration* (Paris, 1996), pp. 49, 110 & 124-126. Daniel Dessert, *La Royale: Vaisseaux et marins du Roi-Soleil* (Paris, 1996), pp. 74-76, 101-110, 170-174, 238-261 & 284-286.

37. Glete, *Navies and Nations I*, 222.

38. I.A.A. Thompson, *War and Government in Habsburg Spain 1560-1620* (London, 1976), pp. 211-283. Somewhat more optimistic views are offered by R.A. Stradling, *The Armada of Flanders: Spanish Maritime Policy and European War, 1568-1668* (Cambridge, 1992), and David Goodman, *Spanish Naval Power, 1589-1665: Reconstruction and Defeat* (Cambridge, 1997).

39. J.S. Wheeler, 'Navy Finance, 1649-1660', *Historical Journal XXXIX* (1996) pp. 457-466.

40. Dessert, *La Royale*; cf. his 'La marine royale, une filiale Colbert', in *Patronages et clientelismes 1550-1750 (France, Angleterre, Espagne, Italie)* ed. Charles Giry Deloison & Roger Mettam (Lille & London, n.d.) pp. 69-83. Martin van Creveld, *Technology and War from 2000 B.C. to the Present* (New York, 2nd edn 1991), pp. 317-319.

41. Verge-Franceschi, *La marine française*. Acerra & Zysberg, *L'essor des marines de guerres*. N.A.M. Rodger, 'The Continental Commitment in the Eighteenth Century', in *War, Strategy and International Politics: Essays in Honour of Sir Michael Howard*, ed. Lawrence Freedman, Paul Hayes & Robert O'Neill (Oxford, 1992) pp. 39-55, at pp. 47-50.

42. On French sea officers see the works of Michel Verge-Franceschi, in this case especially *Marine et Éducation sous l'Ancien Régime* (Paris, 1991).

43. Glete, *Navies and Nations I*, 207-208.

44. The dualist contrast presented by Tilly in 'War Making and State Making as Organized Crime' (p.170), between the 'coercive and self-seeking entrepreneurs' who made wars and states, and the heroic (by implication, pacifist) ordinary people who resisted them and in some cases succeeded in preserving representative institutions, seems to me to bear no relation to rea-

lity, and is admitted by its author to be supported by 'no evidence worthy of the name'. His *Coercion, Capital and European States*, however, takes a more plausible line.

45. Kenneth R. Andrews, *Elizabethan Privateering: English Privateering during the Spanish War 1585-1603* (Cambridge, 1964), pp. 233-235. cf Paul M. Kennedy, *The Rise and Fall of British Naval Mastery* (2nd edn 1983), pp. 22-24.

46. Andrews, *Elizabethan Privateering*, p. 238.

47. N.A.M. Rodger, 'Officers, Gentlemen and their Education, 1793-1860', in *Les Empires en Guerre et Paix, 1793-1860* ed. Edward Freeman (Vincennes, Service Historique de la Marine, 1990) pp. 139-151.

48. James Pritchard, *Louis XV's Navy, 1748-1762: A Study of Organization and Administration* (Kingston, Ontario & Montreal, 1987) pp. 186-205. Henri Legohérel, *Les Trésoriers généraux de la Marine (1517-1788)* (Paris, 1965). Dessert, *La Royale*, pp. 54-77. Acerra & Zysberg, *L'essor des marines de guerres*, pp. 248-250.

49. James Scott Wheeler, *The Making of a World Power: War and the Military Revolution in Seventeenth-Century England* (Stroud, 1999). Patrick O'Brien, 'Fiscal Exceptionalism: Great Britain and its European Rivals. From Civil War to Triumph at Trafalgar and Waterloo', in *The Political Economy of British Historical Experience, 1688-1914* ed. Donald Winch & P.K. O'Brien (Oxford, 2002). Cf Philip Harling & Peter Mandler, 'From «Fiscal-Military» State to Laissez-Faire State, 1760-1850', *Journal of British Studies* XXXII (1993), pp. 44-70.

50. Anderson, *Lineages of the Absolutist State*, p. 135. P.J. Cain & A.G. Hopkins, *British Imperialism: Innovation and Expansion 1688-1914* (London, 1993), pp. 64-89.

51. Rodger, *Safeguard of the Sea*, pp. 257-260, 327-329 & 340-341.

52. This is the central argument of Peter Padfield, *Maritime Supremacy and the Opening of the Western Mind: Naval Campaigns that shaped the Modern World, 1588-1782* (London, 1999).

53. Glete, *Navies and Nations* I, 13. André Corvisier, 'Armées, état et administration dans les temps modernes', *Francia* IX (1980) pp. 509-519, at p. 509. Vicens Vives, 'Estructura administrativa estatal', p. 361.

54. William H. McNeill, *The Pursuit of Power: Technology, Armed Force and Society since A.D.1000* (Oxford, 1983) pp. 113-116. Paul Kennedy, *The Rise and Fall of the Great Powers: Economic Change and Military Conflict from 1500 to 2000* (London, 1988), pp. 21-29. Jeremy Black, *War and the World: Military Power and the Fate of Continents, 1450-2000* (London, 1998), p. 4.

55. Paul Langford, *A Polite and Commercial People: England, 1727-1783* (Oxford, 1989) p. 697.

56. Roughly the argument of Otto Hintze, for whom see Anderson, *Lineages of the Absolutist State*, pp. 134-135; and Ertman, 'The Sinews of Power and European State-Building Theory', pp. 34-35.

# The Economics of Sea Power. Property rights arrangements and institutional change in Elizabethan England

*Nicholas Kyriazis\**  
*Michel S. Zouboulakis\*\**

## *Abstract*

*This article focuses on the institutional conditions that helped England to be transformed into a sea power during the reign of Queen Elisabeth I (1558-1603). Institutional change was the result of a successful coordination of public choices in order to organize national defense and privateering expeditions against Spain's trade and its New World colonies. The creation of the stock-market and the establishment of joint stock overseas trade companies were both the outcome of a large alliance between public and private interests, leading to an effective arrangement of property rights and of the appropriate incentives related to them, inaugurating a path responsible for the development of Great Britain over the next centuries. From a methodological point of view our analysis provides an assessment of the appropriateness of the conceptual framework of property rights in understanding commercial capitalism.*

---

\* Assistant Professor of Economics, Department of Economics, University of Thessaly.

\*\* Associate Professor of Economics, Department of Economics, University of Thessaly.

## 1. The historical background: A Kingdom in Peril

England's situation at the coronation of Queen Elizabeth the First, in 1558, was precarious. It has rightly been argued that during the course of history, the sea has been used as a way of invasion, and just as often as a defensive moat<sup>1</sup>. Concerning the British Isles, the sea was the way of invasion for the Romans, the Saxons, the Scandinavian Vikings, the Danes, the Norwegians of Harold Hadrada and the Normans of William the Conqueror. It looked likely that history would repeat itself through a Spanish invasion during the reign of Elizabeth.

Indeed, after the peace of Chateau-Cambresis of 1559 with the French, England's main enemy was the Spanish - Habsburgian Empire of Phillip the Second (1556-1598). It was a multinational empire of great geographic extent, encompassing Spain, the Low Countries, great parts of Italy in the North and South, Sicily and Sardinia, Portugal (annexed in 1580), Franche Comte, and the vast empire in Central and South America as well as the Philippines in Asia. The population of the European part of the empire was about 20 millions, of which 10 million were in the Iberian Peninsula, out of a total of 95 million for all of Europe including the European parts of Russia and the Ottoman Empire<sup>2</sup>.

In 1580, England and Wales had a total population of about 4.5 million, and although it had a homogenous population, it could not be considered a great economic or military power. England did not have any colonies and although some of the Channel ports had an important seafaring tradition, its commercial navy was smaller than Holland's, Spain's or France's and comparable to that of Venice or the German Hanseatic cities, both of which had smaller populations. The land armed forces of the empire were comprised of about 200,000, while those of England 30,000<sup>3</sup>. The Spanish army with its famous «tercios» regiments was the best in the world, while its navy had gained resounding victories against the Turks in Lepanto (1571) and the French - Portuguese in the Azores Sea (1580).

The deteriorating political relations with the Spanish empire made a sea born invasion a real menace. England's leaders understood that in order to face successfully Spain's challenge, they had to adapt their military strategy and become a naval power, just like Athens did under similar circumstances, when faced with the Persian

invasion of 480 BC. Two major improvements were introduced in naval technology, namely galleons and iron guns.

During the second half of the 16th Century, the galley, a direct descendant of the ancient Greek «trieris», was the dominant warship type in the Mediterranean. It was totally unsuitable due to the need for greater range and for facing big waves, in the open Atlantic waters<sup>4</sup>. So, new ship types had evolved, the naves, carracks and caravelles, which had sails as their sole propulsive power. For military purposes, they evolved into the galleon, a long ship with clean lines, armed with guns of various calibers, on usually two levels at their sides. The fire of a galleon's broadside could sink any galley. During the period under consideration, neither England nor Spain had built enough purpose galleons, so that both fleets pressed into service armed merchantmen. Although Spanish and English ships had more or less the same displacement, English galleons were faster and more maneuverable and had a marked superiority in their rate of fire, due to more efficient gun carriages, that allowed faster reloading<sup>5</sup>.

In the field of artillery, England, forced by circumstances, made substantial progress by 1540. Till then, most guns were made of bronze. In the 1540s, the first iron foundry was established with the Crown's support, at Weald in Kent. In 1543 an «international» team with Frenchman Pierre Baude and Englishmen William Levett and Ralph Hogg produced the first cast iron guns. From a technological stand point, iron guns had many disadvantages compared to bronze ones,<sup>6</sup> but they offered one decisive advantage for England: Iron abounded in the country and cost one tenth of the price of bronze. During a period of increasing inflation, the price of cast iron guns decreased from 10-12 pounds per tone in 1565-1570, to 8-9 pounds in 1600<sup>7</sup>.

The necessity of creating a «defense industry» as it is called today, is a classic example as to how technological progress in this sector creates links and spillovers that are beneficial to the growth of the economy as a whole<sup>8</sup>. During a period of forty years, the iron industry was so developed that in 1573 there existed seven foundries producing 400 tons of guns and shot per year, being only part of the total iron industry. In 1520 there existed just one blast furnace, increasing to 67 with a capacity of 1300 tons by 1570<sup>9</sup>. So, England, which by the beginning of the century was an importer of guns, had become the foremost producer and exporter of cheap iron guns by the end of the century, being practically the

only country to possess that technology<sup>10</sup>. Although it is broadly accepted that the industrial revolution in England began at about 1750, it is clear from the above, that as far as the iron industry is concerned the process characterized by mass production, very high rates of growth and import substitution by exports, began during the second half of the 16th Century.

Hence, pressed by necessity, England developed an improved weapon's system that consisted of the new fast galleon armed with iron guns on efficient four-wheel carriages. This partial naval superiority would not be sufficient, if it was not combined with a pertinent improvement of strategy and tactics. The Spanish plan was based on collaboration of naval and land forces. The fleet under the Duke of Medina-Sidonia would sail from Spain to join the army of the Duke of Parma at Calais, which was 60,000 strong. There, the army would embark and the fleet would transport them for a landing on the southern English coast, in a repetition of William the Conqueror's landing<sup>11</sup>. If it succeeded in this, the subjugation of England appeared almost a foregone conclusion, because the English did not have an army strong enough to defeat the Spanish. The weak point in the Spanish strategic plan was that not possessing in Spain itself the necessary army to embark directly on the fleet, they had first to join forces with Parma, before attempting the invasion<sup>12</sup>. Success depended on land and sea forces coordination and the prevention of their meeting became the focus of English strategy.

The Spanish based their naval tactics on boarding the enemy's vessels and capturing them. So they embarked a strong element of naval soldiers. The English wanted to avoid boarding and transform the battle into a long-range gunnery contest. Having faster and more maneuverable ships and a faster rate of fire, they managed to achieve this, but they did not succeed in breaking the Armada's admirable discipline and defensive order. The Armada anchored before Calais, awaited Parma's embarkation. At this crucial moment, at dawn of the 7th August 1588, the English used their secret weapon, fire ships. Although the Spanish managed to avoid them, the attack achieved its aim. The Spanish ships had to cut and leave their anchors, in order to sail in a hurry. But without anchors they could not come back to embark the army and they had to continue sailing on, contouring the British isles and loosing 42 out of their 107 ships in storms and shipwrecks<sup>13</sup>.



## 2. Political organization and public finance

The transition of a country from a land to a sea power demands structural organizational modifications and results in significant social changes. Such was the case of Athens's transformation after the Persian wars, and England followed the same rule. For the Armada campaign, the English fleet comprised 110 warships (galleons and converted merchantmen), 15 supply ships, 6 small warships of less than 100 tons each plus 95 other small ships for various duties. The total crews of the ships have been estimated at about 16,000<sup>14</sup>. Naval population in England was more or less stable for the period 1500-1550, but demand for manning the royal ships increased, reaching a peak in the Armada years, when out of a total naval population of 16,255 in 1583 the fleet needed about 16,000<sup>15</sup>. So, even if all English sailors were employed in the fleet, their total number, taking into account illnesses, desertions etc., would still be insufficient. The demand gap had to be covered from other sources of the population, these being the inland counties. As early as in 1586 these counties were ordered to supply men (farm laborers, workers, and artisans even condemned criminals) to man the fleet. During the last twenty years of Elizabeth's reign (1583-1603) an increasing number of inhabitants of inland counties would abandon the difficult living conditions of the countryside to seek their fortune in the sea. The very lucrative voyage of Drake's ship «Golden Hind» in 1577 had fired the fantasy of the courageous young Englishmen, turning them to the sea. As a result, England's naval population, after remaining stable for about the first 60 years of the 16th century, had at least doubled during the last twenty.

Opening itself to the sea, England not only laid the foundations of sea power, but also transformed its political system towards increased freedom and democracy, a change that would become more apparent during the next century. Naval population was clearly more progressive than the purely landlocked one and this again manifested itself during the English Civil War. The naval cities and districts supported Parliament against the Stuarts, contributing substantially to its victory<sup>16</sup>. Once again in history, naval power and constitutional government went together. Strong navy requires great expenses, only affordable by strongly supported and stable governments capable of translating that support into longterm political commitment<sup>17</sup>.

A thorough comparison of the economies of the Spanish Empire and England seems on a first look, strongly in favor of the Empire,

which had five main sources of revenue: 1. Direct taxes like the tax revenue of the Spanish Kingdom, mainly Castile; other taxes like that paid by the «Mesta» Guild of shepherds producing wool, sales tax «tomsa», and «crusading tax» on ecclesiastic property. 2. Taxes levied in Italy and the Low Countries. 3. Custom duties levied in Italy and the Low Countries. 4. The American colonies, mainly from the silver Potosi mines in Peru (discovered in 1543). 5. Profits from bank capital invested in the most important banking centers of the time belonging to the Empire, namely Milan, Seville, Cadiz and mostly Antwerp.

But on the other hand, public expenditure increased even faster during the 16th Century, because the Empire fought continuously on many fronts. In the Mediterranean and North Africa against the Ottoman Turks and their tributary corsair kingdoms of Algiers and Tunis, in Italy against the French and some of the Italian states, in Germany and Bohemia supporting the German Habsburg branch against the Hussites and the German Protestant Princes, in the Low Countries after 1572 against the rebelling Dutch and during the last twenty years also against the English. Spain was suffering from «imperial over-stretching» not being able to, despite its strength and wealth, overcome all its enemies at the same time<sup>18</sup>.

The cost of war made obvious the financial weakness of Spain. When Charles V resigned in favor of his son Phillip II in 1556, the latter inherited a public debt of about 20 million ducats, which could not be served by Spain's normal revenues, resulting in Spain's bankruptcy the following year. This brought down also its debtors, especially the great bankers from Augsburg, the Fugger family<sup>19</sup>. Phillip's revenues doubled during the period 1556-1573 and more than doubled, to the end of his reign. The revenues from America's silver mines have been estimated at 12 million ducats for 1571-1580 and over 25 million for 1591-1600, but expenditure rose even faster. The Armada's cost was 10 million ducats in 1588 and Phillip had serious problems in financing his fleet. So, he was obliged to recur to increasing loans mainly from the Genoese bankers under increasingly negative conditions: The Genoese gave cash loans to Phillip in exchange for contracts, known as «asientos», which tied to them increasing future quantities of silver from the American silver mines<sup>20</sup>. In 1596, when Spain went bankrupt for the third time within the same century, public debt had risen above 100 million ducats, so that interest rates on it took two thirds of the total revenue<sup>21</sup>. Within twenty years, Phillip's debt increased four fold<sup>22</sup>.

A dramatic economic consequence of the war effort was the «great inflation» resulting from the inflow of precious metals, mainly American silver, without an analogous increase in production. Nominal agricultural prices rose by 500% and «industrial» ones by 300% between 1500 and 1630. Scholars living at that time, like Salamanca's Professor Martin de Azpilcueta, noticed with surprise this phenomenon, trying to give various explanations. This phenomenon gave rise to a passionate debate between Sieur de Malestroit who suggested that depreciation of money was due to an internal devaluation of the currency and Jean Bodin who has tried to explain «inflation» as a result of the great increase of currency circulation (money supply), in a crude presentation of the quantity theory of money<sup>23</sup>.

When Elizabeth was crowned in 1558, after the death of her sister Mary Tudor, first wife of Philip, she brought to Parliament a naval law-budget, for the construction of «the strongest wall of defense that can exist against the enemies of this island»<sup>24</sup>. A permanent naval budget was thus introduced, under the supervision of the Treasurer of the Navy, which enabled England to lay down the nucleus of its fleet of war galleons, 14 of which were built in the five years till 1564. The naval program gave an impetus to the steel industry, as stated above, and to the shipbuilding industry. During 1548-1551 only 100 persons were working as paid members by the English Treasury in the dock yards, while already by 1559, 520 «shipbuilders» and 100 «workmen» worked at the three royal shipyards at Deptford, Woolwich and Portsmouth. By comparison, the greatest «private industry» of the time was a turf producing plant in London, which employed 7 workers and 11 apprentices while the biggest «industrial firm» had a total capital of 20 pounds!<sup>25</sup>

Elizabeth's public revenue in 1575 was 300,000 pounds (at an exchange rate of five shilling and six pence per ducat), only one-eleventh of Phillip's revenue. After ten years of good administration, by 1585 Elizabeth had a surplus of 300,000 pounds. As soon as hostilities with Spain began, the Parliament voted an additional expenditure of 72,000 pounds per year. At the beginning of 1588, Elizabeth, thanks to the sound financial advises of Sir Thomas Gresham, had a surplus of 154,000 pounds. In the same year the English fleet cost over 150,000 pounds, while for the period 1580-1590 total yearly average public expenditure rose above 350,000 pounds<sup>26</sup>. The English Crown tried to cover a part of this expenditure by selling some of the Crown's land, or monopolies revenues.

Revenues arising from those two sources were insufficient, so that Elizabeth had to convene the Commons and ask for additional taxation and grants for a total of 2 million pounds. When the Commons voted favorably, the Crown was in a position both to pay the sailors of the fleet and to avoid bankruptcy, while Phillip often could not achieve the first and did not avoid the second. But, on the other hand, the fact that the Crown had to summon the Commons resulted in strengthening its position and increasing its political power as well as of the classes and regions that it represented.

### 3. Property rights' structure and economic efficiency

The Anglo-Spanish War is a very good example of how a smaller country, when it manages to create more efficient economic institutions, can overcome a stronger military and economic power. As many historians have suggested<sup>27</sup>, Phillip's Empire was a «model» of inefficient and short-viewed economic policy. The tax system was unjust, weighing heavily on particular regions and classes, giving rise to counterincentives and distortions. Castile bore the main tax burden. A representative body, the «Cortes» existed also there, but with much reduced power to control public expenditure and with the important difference as compared to the Commons, that in the Cortes were represented almost solely the landed aristocracy without also the commercial, industrial and financial interests of the country. Taking this into account, it is not strange that the Cortes approved taxes that had to be paid by the other classes and not the aristocratic landowners, such as the «alcabala» (10% levy on commercial exchanges), the «encabeziamento» (a yearly per head tax), the «millones» (a tax on alimentation products), the «servicios» (grants) etc.<sup>28</sup> All these taxes, together with the internal custom duties between the various ex-Kingdoms (Castile, Aragon, Leon, Andalusia, Luzitania etc) of the Spanish peninsula inhibited commerce and resulted in a reduction of the incomes of the poorer classes of the population, while at the same time discouraging the initiative for new productive activities. Not accidentally, the «hidalgos» distrusted commerce and chose a career in the church or the Army, the only institutions that the Crown did not dare to touch<sup>29</sup>. The system of counterincentives resulted in a deficit of specialized labor

and capital for investment purposes which had its direct repercussion also in the military effort: In contrast to England, in Spain itself there was no development of a gun producing and iron industry except for a few small foundries, and «Spanish» guns were mainly foreign imports from Italy, the Netherlands and Germany, bought with silver from the New World.

A second example of the counter-productive structure of property rights, is the prerogatives of the «Mesta» guild of sheep owners and shepherds, which the Crown had already granted them in the 13th century. The «Mesta» was taxed for the right to graze its sheep. In exchange, the shepherds and their flocks could roam the countryside and feed anywhere, without any prohibitions. This led practically to the inexistence of property rights in land that could be used for extensive high quality, high yield agricultural products. As a consequence, Spain was all the more dependent on foodstuff imports, mainly cereals for its increasing population, with frequent outbursts of famine. Moreover, not having its own merchant fleet, these imports were being made by using ships of other countries, (Dutch, French, English, Genoese, Venetian) which were not infrequently in a state of war with Spain. The directive to the shipyards of the Biscayan ports that they should build warships as against commercial vessels, especially after the Armada's losses – and the custom duties on wool exports which benefited the English and Flemish competitors<sup>30</sup>, aggravated the inefficiency of the economic system. Spanish products became even less competitive in international markets, as an early manifestation of the phenomenon that has been recently called «Dutch disease»<sup>31</sup>.

On the other hand, Elizabeth's government managed to stabilize the pound and control inflation thanks to Gresham's monetary reform. During the war years against Spain, thanks to a cautious and efficient fiscal policy, it managed to spend only 4.5 million pounds. This corresponded to one third of the total public expenditure<sup>32</sup>. The basic difference with Spain was that England covered the expenses mainly from its yearly revenues, avoiding external loans and eventual bankruptcy. In parallel, England's armed forces were regularly paid, in opposition to those of Spain which were not. Also, at the initiative of the English Admiralty, an innovative system of assistance for sailors and their families was introduced which took over their economic support and health care in the case of maiming, illness or death.

However, the true institutional change came out of the reorganiza-

tion of property rights in overseas commerce. The English rulers realized that the main source of Spain's power and wealth were the American colonies and their silver and goldmines. So, they decided to hit and if possibly destroy, not directly the Habsburg's armed forces, but the economic foundations that financed them, inaugurating an early form of «total war». The legal status of trade with the Spanish-American colonies was conducive to interference. From the Treaty of Tordesillas of 1494 between the Spanish and the Portuguese, the New World was divided in two spheres with the Western Indies going to Spain and Brazil, Africa and Asia being given to Portugal, to the exclusion of all other European countries. The Spanish introduced a severe monopoly of trade with their colonies, excluding all foreign traders and considering them as pirates. So, all non-Spanish merchants wishing to trade with the Spanish colonies had to arm their ships with guns for defensive purposes and always be ready for combat. From this point on to open piracy the step was a small one. On the other hand, in many cases, Spanish settlers themselves put the Spanish monopoly aside and traded with French and English merchants buying mainly African slaves (whom the latter had bought from the Portuguese), and selling them in exchange for sugar, tobacco, pearls and skins.

English, French and Dutch attacks against the Spanish New World settlements had already started during the reign of Queen Mary Tudor (1553-1558), sometimes as purely piratical enterprises, and sometimes, when a state of war between England and Spain existed officially. From 1557 on as corsairs whereby every ship and its captain was provided with a «letter of marque» or a «letter of reprisal», which permitted to the said ship to attack the ships and settlements of the enemy. The ship and its captain were no more an outlaw pirate but a corsair or privateer, i.e. a private person in the service of his Queen and country, officially empowered to attack the enemy's trade<sup>33</sup>. Of particular importance for England was the fact that these enterprises during the second half of the 16th century linked trade, war, piracy and privateering with the religious feeling of «holy» war of the Protestants against the Catholic League, with patriotism and economic gains. These gains were distributed from the highest levels of society, the Queen and the Lords of the Admiralty (who received 10% of the sales at auctions of the seized enemy property), to the ship's captains, to the financiers-bankers of the expeditions, to the merchants who bought and resold these goods, as well as to the common sailors and soldiers who had their share in profits<sup>34</sup>.

While at the beginning of the period privateering had much less importance than other economic activities, such as the wool industry, by the end of the period the great change had been accomplished. Due to England's successful response to Spain's invasion threat, England turned itself to the sea, and the «naval economy» with its linkages with other sectors (such as the iron industry, shipyards etc) had become by far the most important. Furthermore, this «turn to the sea» led to the establishment of a new institutional framework concerning property rights that was conducive to sustained economic growth. In a period when government was entrusted to the local gentry and to landowners who were not paid by the crown for their services, Elizabeth did not have the power and probably neither the will to stop piracy during peace, even if such activities were diplomatically harmful. But in times of war the pirates were immediately transformed into privateers collecting a valuable pool of experienced sailors and warriors. Already by 1563 Elizabeth granted the first «letters of reprisal». During the next twenty years privateers would enrich themselves together with their crews and investors becoming at the same time famous national heroes.

Most famous amongst them was one of the Armada's victors, Sir Francis Drake, who undertook his first raid against Spain's New World possessions in 1571. In 1577 he repeated his expedition with his ship «Golden Hind», which would result in the second world circumnavigation voyage after Magellan's. His investors were nobles, members of the Royal Council like Sir Francis Walsingham, the Secretary of State, Queen Elizabeth herself and London banking circles. In the Pacific Ocean, outside the coast of Peru, he captured the Spanish galleon *Nuestra Senora de la Concepcion* loaded with 26 tons of silver. At the Molucas islands he added a valuable cargo of spices, before returning to Plymouth in 1580. The value of his plunder was 600,000 pounds, double the annual Crown revenue, while his investors had a rate of return on their investment of 4,700%! It is clear that such achievements turned the ambition of every adventurous young Englishman, as well as of every «professional» investor, merchant, banker, or nobleman to the sea.

Attacks were not restricted to enemy vessels, but were made also against sea towns. The expeditions increased continuously in the number of men and of ships. If at the beginning of the 1570's the norm was single ships, in the 1590's «private» fleets undertook the expeditions including as many as 26 ships in the last joint expedition of Drake and Hawking against Puerto Rico in 1596. Historians

estimate that the cargoes of seized Spanish ships valued 100,000-200,000 pounds per year for the last two decades of the 16th century, or about 10-15% of total English imports<sup>35</sup>. During the first three months of the 1587-1588 war, English privateers took 27 Spanish ships for a total value of 294,500 ducats. During the whole course of the war, they took over 1,000 Spanish ships. The English organized a total of 150 expeditions against the Spanish New World possessions, leading to the conquest and plunder of many Spanish towns and settlements in the Americas, such as Puerto de Caballos in Mexico (plundered six times), Porto Bello in Panama, even Cadiz in Spain in 1587 and 1596<sup>36</sup>. The successes were such that the prices of colonial goods of the Spanish empire were often cheaper in the market of London than that of Seville!

The privateering expeditions illustrate dramatically what is actually called PPP (Public Private Partnership) concerning not only finance but also the practical operational field<sup>37</sup>. Royal ships participated in the privateering fleets, guns from the royal ships, as well sails, cords, ropes etc were used to furnish «private» merchant ships and corsairs, the captains of which were often officers of the Royal Navy. The final outcome of all those mixed commercial-military enterprises was the establishment of England's, and later on Great Britain's, political and economic dominance till the beginning of the 20th century. England's transformation, by the end of 16th century, into a worldwide naval power started from the overseas «trade» of those mixed commercial-privateer enterprises. All together, as we will see below, Metropolitan England developed the economic institutions necessary for growth. As Rodger remarks «twenty years of war changed England's perspective. While before the Queen possessed an imposing but almost isolated fleet (of war galleons) now naval strength had become a national matter... It was clear that England's future laid in the open seas... The English had learned that the sea was more than a defense against a hostile world: it had become the means to discover new worlds for gold, fame and glory»<sup>38</sup>.

#### 4. Institutional arrangements and economic growth

Political institutions in a representative parliamentary democracy provide a structure of multiple veto points that entail the uncontrolla-



ble activity of central authority, creating a system of «check and balances»<sup>39</sup>. In Elizabethan England the feasibility of economic growth was based on a legal framework that protected efficiently property rights from abuse by the central power. Instead of exercising tight control on property rights and taxation like in Spain, the English crown opted for elastic control and granting of substantial rights first to the privateers-entrepreneurs and then to merchants-entrepreneurs in the wool industry.

Long-run successful economic performance of the English economy was furthermore based in the formation of «correct behavioral incentives» not only for economic but also for political agents. The English crown showed a long-run dedication to the safeguard of this framework, avoiding changing it in its favor according to circumstances<sup>40</sup>. The Spanish crown on the contrary, followed an arbitrary policy without an appropriate institutional basis that would bind it in the future. As a result, six consecutive bankruptcies from 1557 to 1647, which brought down with them great financial houses, ruining its financial credibility, destroyed also in the long-run the eventual development of a financial and banking system in Spain. It is not by chance that during the period of Spain's hegemony, all financial centers were located in regions out of mainland Spain and even out of its jurisdiction, like Augsburg, Antwerp, Genoa, Florence and last of all Amsterdam. England not only avoided bankruptcy but continued paying back Henry VIII's debt, thus laying down the foundations of the financial / banking system that made London later the financial center of the world.

North and Weingast (1989) remarked that the war's issue was in doubt and that possibly the crown might have prevailed if England had a standing army loyal to the crown, similar to the one already in existence in France and Spain. Our previous analysis offers an additional answer to the question why such an army did not exist yet in England: its geographic position and Phillip's invasion threat, forced England to opt out for a particular kind of the public good «defense», in the form of a standing navy, instead of a standing army. This choice, as pointed out above, not only gave it victory against Spain, but also laid down the basis for democracy and the institutions that led to sustained growth. By creating a wide «alliance of interests» between sailors, workmen and handworkers of the naval industry, iron industry etc, merchants, entrepreneurs, «capitalists», bankers and noblemen, an alliance that found its expression in Parliament, it gave to it the necessary strength to function as a second

pole of political power against the Stuarts, and thus strengthen the institutions that supported long run stability and sustained growth. The creation of such an «alliance of interests» is one of the most important elements in creating and transforming property rights, more important than vague ideas on justice<sup>41</sup>. When during the next century the successors of Elizabeth, the Stuart dynasty questioned the established institutional system, in particular Parliament's position, where the interests of the upcoming «bourgeois» class were represented, the great Civil War broke out.

What's more, in England, the opening to the sea and the «alliance of interests» became an incentive for the creation of two further fundamental economic institutions: the capital-financial market and the broadly based joint stock trade companies. The London Stock Exchange was built in 1566 with money given by Thomas Gresham on a plot of land belonging to the City of London, in Threadneedle Street. For many years, in the street level of the building were operating commercial shops, and Gresham used the rents paid by these shops to get back the cost of his initial investment. As pointed out emphatically<sup>42</sup> Thomas Gresham was the ideal type of «the businessman who is just as much a public servant as he is businessman and who, though perfectly successful in looking after his own advantage, serves the state in ways that are beyond the competence of the mere public servant».

The coalition of private and public interests is far more important in overseas trade. As seen above, the privateering expeditions were financed also by the crown's direct contribution in collaboration with the financial means of the privateersship owners. A «syndicate of investors» financed Drake's enterprise of 1577, and this was the norm also for further enterprises. In some of them, Elizabeth contributed also her royal warships in addition to her financial contribution. For the 1595 Caribbean enterprise, the Queen contributed two thirds of the total cost of 33,266 pounds. Other main contributors were the Lord Admiral, important nobles of the court and many famous navy captains, such as Hawkins and Raleigh. The bankers and merchants of London took over the management of the company and the furniture of all necessary provisions for the fleet and crews. These enterprises, while aiming at the state's enemy, the Spanish empire, were not part of any long-run strategic aim, other than profit seeking. Profits from each enterprise were financing the next one, in a self-perpetuating process. And profits were very high for everybody, as seen above, from the Crown, the nobles of the

Court, the merchant-bankers (like John Watts whose profit was 32,000 pounds in 1591), the captains and officers of the ships, down to the simple sailors and soldiers.

Before long, the necessary institutional framework was created promoting the development of other sectors. In 1600 was established the famous «East India Company» as a public-private joint stock company with a broad share basis. It became the main instrument of commercial expansion of the British interests in the wider Asia area, till its dissolution in 1857. It followed the examples of similar previous commercial companies (Muscovy Co 1555, Spanish Co 1577, Eastland Baltic Co 1579, Levant Co 1583), which united public and private funds with the purpose of monopolistic exploitation of trade relations with the named geographic regions<sup>43</sup>. These activities generated positive externalities, where the creation of efficient institutions in some sector –the financial one and the overseas trading companies in our case– works as a basis for the development of other sectors of the economy. The Anglo-Spanish war, waged on private economic criteria of organization and for the sake of profit, had so successful long-run strategic implications and so astonishing unintended consequences, that it would have been impossible to achieve otherwise.

## 5. Conclusion: Efficient institutional change and property rights

New-Institutional Economics explains the evolution of economic systems as an attempt to determine and arrange efficiently property rights and to lower transaction costs. For the organization and completion of transactions, resources are needed for the determination and the enforcement of the exchange conditions, for the collection and distribution of information and for the determination and protection of property rights before, during and after the transaction<sup>44</sup>. The economic efficiency of an organization (a firm or a system) and the final outcome (profit maximization or economic growth respectively) depend directly on the way these property rights are determined, especially as far the ownership between the private and the public sector is concerned<sup>45</sup>. In the long run, economic growth is based upon the successful choice of economic institutions that make transactions easier and less costly over long periods of time.

We have attempted to apply this analytical frame to the early period of commercial capitalism. When Elizabethan England faced the Spanish threat of invasion, the particular choice of the public's good defense in the form of a permanent navy, and the creation of a strong fleet helped both to confront the enemy and to obtain naval supremacy. Once this supremacy was obtained, England tried to expand its commercial activities to overcome Dutch hegemony in the seas during the 17th century. The creation of broadly based joint-stock companies aimed at solving the complex problem of organizing and financing big overseas privateering-commercial expeditions. The size of necessary funds was so great, that neither the public nor the private sector alone had the necessary mix of skills and amount of resources to solve it by itself alone. So, these expeditions based on a large «alliance of interests» led to the establishment of a stable legal framework of property rights and political institutions. Since the Anglo-Spanish war lasted for forty years, this choice had the character of a continuous repeated game, so that the «alliance of interests» could be tested again and again, be refined and become permanent, stabilizing furthermore the legal framework and finally leading to «institutional path dependence».

Elizabethan England gave an ingenious solution to the distribution of costs and benefits of the public good defense: those benefiting the most from it, were bearing the heaviest burden. Naval and commercial circles were called to offer their personal capital and labor for the creation and maintenance of sea power. But England went a step further: It transformed war into a self financing commercial private enterprise, with profit as its main incentive, where the question as to who would bear the costs and get the benefits was automatically solved through the creation of joint-stock trade companies. These companies supported England's commerce during the period starting in the early 1600's, and contributed significantly to its future economic growth. Ocean-shipping efficiency grew rapidly after 1600 because of the «decline of piracy and privateering and the development of markets and international trade»<sup>46</sup>. Elizabethan England is an exemplar case of successful organization and institutional change generated by singular historical circumstances that became «locked-in» into a single growth path<sup>47</sup>.

## Notes

1. Rodger N.A.M., *The Safeguard of the Sea*, (London: Harper & Collins 1997).

2. Spooner F.C., «The Economy of Europe 1559-1609», in R.B. Wernham ed. *The New Cambridge Modern History*, vol. III. (Cambridge: Cambridge U.P., 1990).

3. Kennedy P., *The Rise and Fall of the Great Powers*, (Unwin and Hyman 1988).

4. The galleys, being propelled by oars, necessitated big crews and so were dependent on replenishment of water and food at frequent intervals. Furthermore the «mechanics» of oar propulsion necessitated low boards over the sea's surface, which were inadequate against the heavy seas of the Atlantic.

5. Lewis M., *The Spanish Armada*, (Pan, 1960), maintained that English guns had smaller calibers but superior range, a view contested by Rodger *The Safeguard of the Sea*, and Konstam, A. *The Armada Campaign 1588*, (Oxford: Osprey Publishing Co., 2001).

6. These being that iron melts at a much higher level of temperature than bronze, making it thus much more difficult to achieve high quality without shortcomings. It weighed about one fifth more than bronze and became corrugated by gunpowder. A bronze gun that received an over close of powder discharge dilates and cracks, but an iron gun just explodes without further warning. Thus, iron guns were more dangerous to their own crews.

7. Davies C.S.L., *Supply Services of the English Armed Forces*, Oxford, Ph. D. Thesis, (1963).

8. Other examples of this are radar technology, airplane turbines, missile propulsion, computers and atomic energy developed during the second World War for military purposes, and satellites, laser, new composite metals and synthetics developed during the Cold War period. For a critical appraisal of the positive economic effects of war, see Nef J.U., «The impact of war on science and technology», in *4th International Conference of Economic History*, Bloomington 1968, (Paris: Mouton 1973).

9. Cipolla C., *Before the industrial revolution*, (New York, W. Norton & Co. 1980).

10. Schubert H.R., *History of the British Steel Industry from 450 B.C. to AD 1775*, (London, 1957). Cleere H. - Crossley D., *The Iron Industry of the Weald*, (Leicester 1985).

11. Tincey J., *The Armada Campaign 1588*, (Osprey Elite, No 15. 1988), Thompson I.A.A., *War and Government in Habsburg Spain 1560-1620*, (London 1976).

12. 7,667 sailors and 20,459 soldiers manned the Armada. Before his death, the Admiral, Marquis De Santa Cruz estimated that England's con-

quest would necessitate an expeditionary force of 55,000 foot and 1,200 cavalry, to be embarked in 520 sailing and 89 oared vessels for a total cost of 2,589,519 ducats. Phillip found this unachievable and thus selected the option of embarking Parma's army. A propos: Fuller J.F.C., *The Decisive Battles of the Western World*, (Paladium 1970), Rodger N., *op. cit.*, Konstam, *op. cit.*

13. The Armada consisted of 107 big ships, plus 34 smaller ones of less than 100 tons. In total, only 4 big ships were lost from various causes during the Channel naval fights; the rest had sunk in storms. See Konstam, *The Armada Campaign*.

14. Konstam, *op. cit.*

15. Rodger, *op. cit.*

16. The mainly agricultural regions supported the king, while the seaboard «naval» cities and counties, as well as the commercial and «industrial» ones, where sailors, workers, merchants and industrialists predominated, and London -where in addition to the above «classes» there existed also the «bankers» and financiers- sided with Parliament. See Wernham R.B. «The English Question 1559-1569», in Wernham R. (ed.), *The New Cambridge Modern History*, (Cambridge U.P. 1990), vol. III. Weingast, B.R. «The Political Foundations of Limited Government: Parliament and Sovereign Debt in 17th and 18th Century England», in J. Drobak & J. Nye (eds), *The Frontiers of New Institutional Economics*, (San Diego: Academic Press 1997).

17. Glete, J. *Navies and Nations, Warships, Navies and State Building in Europe and America, 1500-1860*, (Stockholm: Coronet Books 1993).

18. Kennedy P., *op. cit.*

19. Lynch J., (1964), (1969), *Spain Under the Habsburgs*, 2 vols, Oxford. Braudel, F., *Civilisation materielle, economie et capitalisme XV-XVIII siècle*, vol III, (Paris: Livre de Poche, 1993).

20. Braudel F., *op. cit.*

21. Koenigsberger H.G., «Western Europe and the Power of Spain», in R.B. Wernham (ed.), *The New Cambridge Modern History*, vol. III. (Cambridge Univ. Press 1990).

22. Braudel F., *La Mediterranee et le monde mediterraneen a l' epoque de Philippe II*, vol. I, (Paris: Armand Colin, 1990).

23. It is nowadays largely accepted that they were both partly right. Cf. Schumpeter J.A.S, *History of Economic Analysis*, (N.Y. Allen & Unwin 1954), pp.311-2; Spooner, *op. cit.*, p. 18.

24. Quoted in Rodger (1997). We have here a striking analogy with the emergence of another naval power, Athens in the 5th century B.C. As soon as the Democratic leader, Themistocles came to power, he introduced his «Naval Law», that enabled Athens to build during 482-481 BC 200 triremes, the fleet that won the sea battle against the Persians in Salamis, 480 BC. See Kyriazis N. - Zouboulakis M., «Democracy, Sea Power and Institutional Change: An Economic Analysis of the Athenian Naval Law», *European Journal of Law and Economics*, 17 (1), 2004.

25. Davies, C.S.L., *op. cit.*
26. Henry VIII's expenditure during 1542-1547 for the navy reached 265,000 pounds out of a yearly revenue of about 250,000. Queen Mary spend 21,000 pounds per year for the Navy, and Elizabeth 16,000 per year for the period 1565-1574, or 6.5% of total «normal» revenues (Rodger 1997). We can compare this percentage with the defense expenditure (total, not just Navy) between 4-6% of GDP for Greece during 1990-2000, 2-2.5% for the EU countries average and 3-3.5% for the USA for the same period.
27. Wallerstein I., *The Modern World System II: Mercantilism and the Consolidation of the European World Economy*, (San Diego: Academic Press 1980), North D.C., *Structure and Change in Economic History*, (London: W.W. Norton & Co. 1981), Kennedy, *op. cit.*
28. Houmanides L., *Economic History of Europe*, (Athens: Papazizis, 1991, in Greek).
29. North D.C., *op. cit.*
30. Kennedy, *op. cit.*
31. After the discovery and exportation of gas by the Netherlands, this has lead to an appreciation of the Dutch currency, thus making all Dutch exports more expensive and less competitive.
32. Braudel F., *op. cit.*, vol. III, p.438.
33. Rodger N., *op. cit.*
34. Konstam A., *Elizabethan Sea Dogs: 1560-1605*, (Oxford: Osprey Publishing Co., 2000).
35. Andrews K.R., *Elizabethan Privateering*, (Cambridge: CUP, 1964).
36. Andrews K.R., *The Spanish Caribbean: Trade and Plunder 1530-1630*, (New Haven, 1978); *Trade, Plunder and Settlement: Maritime Enterprise and the Genesis of the British Empire 1480-1630*, (Cambridge, 1984).
37. Nowadays the issue of PPP in National Defense gains again momentum as in the field of communication and imagery satellites. For example the new British system due to enter into service in 2005 is SKYNET 5, which combines frequencies for civilian and military use, as will be the case also for the Greek Satellite Hellasat, while civilian imagery satellites like Landsat, Spot and Iconos are used also for military purposes, such as mapping.
38. Rodger N., *op. cit.* The last phrase has been attributed to sir Walter Raleigh, England's famous corsair and admiral, who among others organized the first unsuccessful English attempt at settlement in America, in 1589, leaving behind for posterity the name «Virginia» for the region, to honor the «virgin queen» Elizabeth.
39. North D.C. - B.R. Weingast, «Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in 17th Century England», *Journal of Economic History*, vol. 59, 1989, pp. 803-832.
40. North and Weingast in «Constitutions and Commitment» speak about «future's long hand». Such is the incentive a king or sovereign has to obser-

ve his dues and obligations as for example paying back his debts. If he does not do this, he loses credibility and in the future he will find nobody to grant him a loan. A counterincentive may be the sovereign's time preference. If he values his present needs much higher than his future ones, then he may not observe his future obligations, because the present gain of non-observance may be higher than the future loss due to loss of credibility. See Buchanan J. - Brennan G., *Reason of Rules*, (Cambridge: Cambridge UP, 1981), Weingast B.R., «The Political Foundations of Limited Government». The same topic is being analyzed from an ethics and economic efficiency standpoint in Klundert T. van Der, «Economic Efficiency and Ethics», *De Economist*, (1999), No 2.

41. Samuels W.J. (1999), «Theories of Property», *Journal of Economic Issues*, vol. 33, pp. 183-8.

42. Schumpeter J.A.S., *History of Economic Analysis*, p.342.

43. Cameron R. (1989), *A Concise Economic History of the World*, Oxford University Press.

44. Among the huge transaction costs literature, see Coase R., «The Problem of Social Cost», *Journal of Law and Economics*, (1960) vol. 3, pp. 1-44. Coase R., (1992), «The Institutional Structure of Production», *American Economic Review*, vol. 82, pp. 713-719, Williamson O.E., *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*, (New York: The Free Press, 1985), North D.C., «Structure and Performance: The Task of Economic History», *Journal of Economic Literature*, vol.16, (1978), pp. 963-978, North, «Transaction Costs, Institutions and Economic History», *Journal of Theoretical and Institutional Economics*, vol. 140, (1984), pp.7-17; North, *Institutions, Institutional Change and Economic Performance*, (Cambridge University Press, 1990); North D.C., «Economic Performance Through Time», *The American Economic Review*, vol.84, (1994), pp. 359-68.

45. Alchian A. - Demsetz H., «The Property Rights Paradigm», *Journal of Economic History*, Vol. 33, (1973), pp.16-27.

46. North D.C., «Sources of Productivity in Ocean Shipping, 1600-1850», *Journal of Political Economy*, vol. 76, (1968), pp. 953-970. In a more recent study, North (1991) adds that long distance commerce conducted to a series of economic institutional innovations that increased capital mobility, reduced the cost of information and distributed risk more evenly, thus reducing uncertainty.

47. On the meaning of «path dependence», see David P.A., «Clio and the Economics of QWERTY», *The American Economic Review*, vol. 75 (2: 1985), pp. 332-7; Arthur B.W. (1989), «Competing Technologies, Increasing Returns and Lock-In by Historical Events», *Economic Journal*, Vol. 99 (1, 1989), pp.116-131; David, «Why are Institutions the “carriers of history”? Path dependence and the evolution of conventions, organizations and institutions», *Structural Change and Economic Dynamics*, vol. 5, (2, 1994), pp. 205-220.



## Program evaluation methodologies – A comparative assessment

*Konstantinos Aravossis\**

*Efrosini Koutsiana\*\**

### *Abstract*

*This paper presents the methodologies used to execute a program evaluation. Several types of methodologies and techniques on program evaluation can be found in the literature, composing a valuable reference base for the evaluator. By presenting the various methodologies we do not aim to measure their efficiency and propose the best one. On the contrary, we believe that each methodology has some advantages and, moreover, in some cases it could be the most appropriate. The aim of this paper is to summarise the different approaches in evaluation, to gather the various methodologies and to determine the specific circumstances under which each methodology is the most appropriate to follow. A more ambitious aim is to propose some interesting methodology mixes, which can be useful guides when looking for evaluation designs that fit better in the reality of a program.*

### 1. Introduction

In this paper we study the evaluation framework of the public

---

\* PhD in Operational Research, University of Thessaly, School of Engineering, Department of Planning and Regional Development, E-mail: arvis@acci.gr, arvis@otenet.gr.

\*\* MSc in Planning and Regional Development, Email: Ekouts@uth.gr.

programs, which are the programs funded by the government, the European Union or other public organizations. We do not present the methods and techniques used to evaluate an individual's plan or investment. We are going to work on the methods used to evaluate the full coverage of socio-economic programs.

In the framework of European Structural Fund programs for the period 1994-99, evaluation is carried out in three stages (European Commission, 1999):

1. Ex ante evaluation focuses on the planned program. It concerns the coherence and the relevance of the project and the realism of the expected results.
2. Intermediate (or on-going) evaluation focuses on the effects of the very first outputs of the program. It is used by decision makers at all levels. It is complimentary to monitoring and intends to reorient the program mid-way.
3. Ex post evaluation intends primarily to report on the effects of programs.

A method is an ad hoc procedure specially constructed for a given evaluation. The term method is similar to design. In the American literature evaluation design is applied to the way in which a given evaluation is constructed. The terms approach and procedure are similar. They apply to a given method and to the spirit in which this method has been constructed (European Commission, 1999).

In the respective literature an important number of evaluation methods and techniques are proposed. Patton (1986) reports that roughly 132 different techniques of evaluation are in use. The big crowd of evaluation techniques raises the question of their systematic categorisation, so that they can express a unified evaluation value framework and a concrete acceptable theoretical background, which is uniformly used (Lagos & Lianos, 2000).

In the following chapters we present some approaches in program evaluation and the basic program evaluation methodologies. By analysing the approaches we point out the basic principles upon which a method is built upon while, by contrasting them we signalize their main characteristics. Hereafter, the basic evaluation methodologies are described with a critical point of view. We consider the advantages and disadvantages of each method and we proceed in a comparative assessment. Such an assessment indicates in which case it is more appropriate to use each methodology.

Next, we examine the combination of evaluation methods as a

new proposal in evaluation design. In this respect we recognize it as necessary to give some instructions for a successful outcome when combining pure methodologies. For a better understanding of this we explain the basic steps needed to obtain specific methodological combinations.

The study concludes with the evaluation methodologies described and the more flexible approaches proposed.

## 2. Types of evaluation and methodological approaches

### 2.1. Types of evaluation

There are four major approaches in evaluation (European Commission, 1999):

#### a) Experimental Approaches

Experimental approaches treat phenomena of causality without analyzing them directly. The effects of the programs are demonstrated by comparing a treatment group with a control group. Evaluations carried out in this framework are typical of American practices in the 1960s and 1970s. These evaluations are long, cumbersome and not always conclusive.

#### b) Economic Approaches

Economic approaches are based on an individualistic view of society, inspired by welfare economics. From this point of view, the value of a public action is the sum of the benefits it provides to individuals. All individuals are supposed to have a system of references and it is assumed that these systems of reference can be aggregated. These hypotheses are used to give an objective character to evaluations carried out by economic techniques. Despite their elegance, economic techniques have been strongly criticized, in both practical and theoretical terms.

#### c) Naturalistic or Pluralistic Approaches

Naturalistic or pluralistic approaches are based on the idea that the political and social world is a collective construction. This construction results from the interaction of differing social groups that have their own interpretations of the same phenomena and issues.

Consequently, public programs are seen as temporary compromises between groups and actors. By means of appropriate techniques, based to a large extent on group work, evaluation acts as a mediator between the different points of view. It is a tool for promoting conciliation between the stakeholders, and its conclusions are all the more robust when they are a product of consensus.

#### d) Pragmatic Approaches

Pragmatic Approaches adopt simplified views of the processes that they have to describe. Theoretical references are less pure and more eclectic. Evaluation objectives are of essentially managerial rather than scientific nature. Evaluations carried out in this spirit, using techniques inspired by management approaches, are oriented more towards the implementation and improvement of efficiency. Their cognitive dimension is less important than in other forms of evaluation.

## 2.2. Methodological approaches in evaluation

### 2.2.1 *Quantitative & Qualitative approaches*

Quantitative analyses focus on testing hypotheses and use structured designs and statistical methods to analyze data. This type of information needs standardization, precision, objectivity and reliability of measurement (internet).

Qualitative approaches, in contrast, gather data in a more open-ended fashion. Data collection usually occurs in natural settings, and focuses more on experiential or subjective aspects of a program. These data can include narrative accounts and may employ multiple data collection techniques (Worthen, Sanders & Fitzpatrick, 1997). Evaluators use qualitative designs to help them understand and describe program implementation rather than to demonstrate statistically significant effects (internet).

### 2.2.2 *Content & Inductive approaches*

Content approaches involve identifying coherent and important themes and patterns in the data. The analyst looks for quotations or observations that go together. Practically, that means pulling together all the data that address a particular evaluation question and then, subdividing that data into coherent categories, patterns and themes (Patton, 1987).

Labeling the data and establishing a data index are the first steps in the Content Approaches. Subsequently, it is critical to have a classification system for the contents of the data. Organizing and simplifying the complexity of data into some meaningful and manageable themes or categories is the basic purpose of the Content Approaches (Patton, 1995). Generating useful and credible qualitative evaluation data through observation and interviewing, in the frame of content approaches, requires discipline, knowledge, training, practice and hard work (Patton, 1987).

By contrast to Content Approaches, Inductive Approaches mean that the patterns, themes and categories of analysis come from the data. They emerge out of the data rather than being decided prior to data collection and analysis. The analyst looks for natural variation in the data. For evaluators, the study of natural variations will involve particular attention to variations in program processes and the ways in which participants respond to and are affected by programs (Patton, 1995).

### 3. Program evaluation methodologies

#### 3.1. Efficiency analyses

Knowledge of the extent to which programs have been implemented successfully and the degree to which they have the desired outcomes is indispensable to program managers, stakeholders and policymakers. In almost all cases, however, it is just as critical to be informed about how program outcomes compare to their costs (Rossi and Freeman, 1993).

Efficiency assessments –Cost Benefit Analysis and Cost Effectiveness Analysis– provide a frame of reference for relating costs to program results. The procedures employed in both types of analysis are often highly technical and their applications will be described only briefly here (Rossi & Freeman, 1993).

##### 3.1.1 *Cost-Benefit Analysis*<sup>3</sup>

Cost-Benefit Analysis estimates and totals up the equivalent money value of the benefits and costs of projects to the community, in order to establish whether they are worthwhile.

In order to reach a conclusion, as to the desirability of a project,

all aspects of the project, positive and negative, must be expressed in terms of a common unit. The most convenient common unit is money. This means that all benefits and costs of a project should be measured in terms of their equivalent money value. A program may provide benefits which are not directly expressed in terms of euros but there is some amount of money the recipients of the benefits would consider just as good as the project's benefits.

Not only do the benefits and costs of a project have to be expressed in terms of equivalent money value, but they have to be expressed in terms of euros of a particular time. This is not just due to the differences in the value of euros at different times because of inflation. A euro available five years from now is not as good as a euro available now. This is because a euro available now can be invested and earn interest for five years and would be worth more than a euro in five years. If the interest rate is  $r$  then a euro invested for  $t$  years will grow to be  $(1+r)^t$ . Therefore the amount of money that would have to be deposited now so that it would grow to be one euro  $t$  years in the future is  $(1+r)^{-t}$ . This is called the «discounted value» or «present value» of a euro available  $t$  years into the future.

When the euro value of benefits at some time in the future is multiplied by the discounted value of one euro at that time in the future, the result is discounted to present value of that benefit of the project. The same thing applies to costs. The net benefit of the projects is just the sum of the present value of the benefits less the present value of the costs. The choice of the appropriate interest rate to use for the discounting is a separate issue.

If the discounted present value of the benefits exceeds the discounted present value of costs then the project is worthwhile. This is equivalent to the condition that the net benefit must be positive. Another equivalent condition is that the ratio of the present value of the benefits to the present value of the costs must be greater than one.

If there are more than one mutually exclusive projects that have positive net present values then there has to be further analysis. From the set of mutually exclusive projects the one that should be selected is the one with the highest net present value.

Costs are either one-off, or they may be ongoing. Benefits are most often received over time. We build this effect of time into our analysis by calculating a pay-back period. This is the time it takes for the benefits of a change to repay its costs.

The impact of a project in the Cost-Benefit Analysis is the difference between what the situation in the study area would be with and without the project. This means that when a project is being evaluated the analysis must estimate not only what the situation would be with the project but also what it would be without the project. In other words, the alternative to the project must be explicitly specified and considered in the evaluation of the project. Note that the with-and-without comparison (ex ante evaluation) is not the same with the before-and-after comparison (ex post evaluation).

### 3.1.2 *Cost Effectiveness Analysis*

By contrast to Cost Benefit Analysis, Cost Effectiveness Analysis does not require the benefits and costs to be reduced to a common denominator. Instead, the effectiveness of a program in reaching given substantive goals is related to the monetary value of the costs (Levin, 1975).

In Cost Effectiveness Analysis, programs with similar goals are evaluated and the costs compared. Cost Effectiveness Analysis, thus, allows comparison and rank ordering of programs in terms of their costs for reaching given goals, or the various outputs required for different degrees of goal achievement (Rossi and Freeman, 1993).

Cost Effectiveness Analysis is based on the same principles and utilizes the same methods as Cost Benefit Analysis. The assumptions of the method, as well as the procedures required for measuring costs and discounting, for example, are the same for both. Therefore, the concepts and methodology introduced previously with regard to Cost Benefit Analysis can also be regarded as a basis for understanding the Cost Effectiveness Analysis (Rossi and Freeman, 1993).

## 3.2. Impact analysis

All impact assessments are comparative. Determining impact requires comparing, with as much rigor as it is practicable, the conditions of targets that have experienced an intervention with those of equivalent targets who have experienced something else (Rossi and Freeman, 1993).

Full coverage programs present special difficulties to evaluators attempting impact assessments since there are no un-served targets available to use as controls. The only comparisons available to the researcher are between the same targets before and after exposure to

the intervention, which are called reflexive controls (Rossi and Freeman, 1993).

Although few evaluation designs have as much intuitive appeal as simple before and after studies, they are among the least valid of assessments. The essential feature of this assessment is a comparison of the same targets at two points in time, separated by a period of participation in a program. The differences of the two measurements are taken as an estimate of the net effects of the intervention (Rossi and Freeman, 1993).

Another applicable design in the course of impact assessments is the Time Series Analysis. The Time Series Analysis involves many repeated measures. The measures are taken on an aggregate unit with many data points preceding and following the point in time at which a new full-coverage intervention was introduced or an old program was substantially modified. By «aggregate» statistical series, we mean periodic measurements taken on a relatively large population (or parallel samples of it) as, for example, vital statistical series (births, deaths, migrations) (Rossi and Freeman, 1993).

Time Series Analysis is especially important for estimating the net impacts of changes in full coverage programs, particularly those that are delivered uniformly. For example, social security retirement payments are uniform for all persons with the same pre-retirement employment records. If retirement payments or sanctions for convicted felons are changed at some point of time then the impact of those changes can be studied through Time Series Analysis (Rossi and Freeman, 1993).

### 3.3. Planning balanced sheet

Planning Balanced Sheet Analysis uses monetary units for the measurements. Time dimension and physical scales are added in the assessment while equity principles are incorporated (KEPE, 1997).

This analysis entitles the goals to be measured as a mirage of the preferences of the agents being surveyed. A main disadvantage is that a person is assumed to belong to only one group, while his other attributes may be omitted (KEPE, 1997).

### 3.4. Goal achievement matrix

Goal Achievement Matrix consists in evaluating a program by the extent of the objective goal's achievement. Like Planning Balan-



ce Sheet, Goal Achievement Matrix uses monetary units for the measurements. The novelty in this assessment is that the evaluator enjoys the benefits of flexibility to examine each goal through several points of view due to the form of the Matrix (KEPE, 1997).

### 3.5. Multicriteria analysis

Multicriteria Analysis takes into account the conflicts and reconciliations between all involved interested parties (KEPE, 1997).

Nijkamp (1997) inducts the Concordance Analysis, a three phase process, for the confrontation of problems in Regional Planning. The process is summarized in the following steps: Some criteria are determined, grouped and compared per two. Then they are presented in a matrix in order to be calculated with indicators of concordance or discordance. Nijkamp considers that, thereafter, Cost Benefit Analysis may be used in the calculating part of the assessment (KEPE, 1997).

Holmes (1972) proposes a process of ranking. The ranks of goals and criteria are determined in advance. The innovations that gather more ranks are the most prospective to be chosen. The major disadvantage of Ordinal Ranking methods is that they comprise subjective elements (KEPE, 1997).

## 4. A comparative assessment

In this chapter we attempt a comparative assessment of the evaluation approaches and methodologies reported above. Firstly, the approaches and methodologies are criticized separately. Following this, the methodologies are presented in a table and their main characteristics are compared. Furthermore, the methodologies are assessed in regard to the approach that they involve.

### 4.1. Comparative assessment of approaches

#### 4.1.1 *Quantitative & Qualitative approaches*

A qualitative evaluation design might be particularly appropriate where, for whatever reasons, either program processes or program impacts, or both, were largely unspecified. Sometimes the reason is

because outcomes were meant to be individualized; sometimes the program is simply uncertain about what the outcomes will be. Under such conditions one purpose of the evaluation may be to help articulate program processes, program impacts and the linkages between the two (Patton, 1995).

Qualitative methods permit the evaluator to study selected issues, cases or events in depth and detail. The fact that data collection is not constrained by predetermined categories of analysis contributes to the depth and detail of qualitative data. Quantitative methods, on the other hand, use standardized measures that fit various options and experiences into predetermined response categories. The advantage of the quantitative approach is that it measures the reactions of a great number of people to a limited set of questions, thus facilitating comparison and statistical aggregation of the data. This provides a set of findings, which are broad and applicable for generalization. By contrast, qualitative methods typically produce a wealth of detailed data about a much smaller number of people and cases (Patton, 1987).

An unfortunate debate has arisen about which of these two approaches is «best.» This debate is not productive. A combination of approaches frequently yields the most useful information. For example, when first approaching an evaluation, an open-ended approach can provide information that will assist the evaluator in developing more quantitative measures. Qualitative measures will provide illustrations and examples that stakeholders may find helpful in understanding the effect of the program on individuals. Further, stakeholders, and members of the general public tend to remember and be moved by case illustrations that describe an individual's experience (internet).

#### 4.1.2 *Inductive and deductive approaches*

In evaluations, the classic deductive approaches are measuring relative attainment of predetermined clear, specific and measurable goals. By contrast, the classic inductive approaches are goal free evaluations in which the evaluator gathers qualitative data on actual program impacts through direct observations of program activities and in depth interviews with participants, without being limited to stated, predetermined goals (Patton, 1987).

## 4.2. Comparative assessment of methodologies

### 4.2.1 *Efficiency Analyses*

The employment of Efficiency techniques (Cost Benefit and Cost Effectiveness Analysis) is appropriate for all phases of program implementation. However, Efficiency Analyses are most commonly undertaken either during the planning and design phase of an initiative, or after an innovative or markedly modified program has been in place for a time and there is interest in making it permanent or possibly expanding it (Rossi and Freeman, 1993).

Ex ante Cost Benefit Analysis is most important for those programs that will be difficult to abandon once they have been put into place, or that require extensive commitments in funding and time to be realized. Decisions on the application of technology and science to public health, medical care and environmental problems are preceded in many cases by ex ante Cost Benefit Analysis because of the extensive resource commitments required. But, most commonly, in the social program field, Efficiency Analysis takes place after the completion of an impact evaluation, when the net impact of a program is known. The focus of such ex post Cost Benefit and Cost Effectiveness assessments may be on examining the efficiency of a program in either absolute or comparative terms, or both. In all cases, the analysis is undertaken to assess whether the costs of the intervention can be justified by the magnitude of the net outcomes (Rossi and Freeman, 1993).

#### 4.2.1.1 *Cost-Benefit Analysis*

In this approach, very little attention is paid to the problems of finding and setting parameters. Analysts tend to include certain benefits and costs in their research because others have done so in the past. Thus each sort of program develops its own conventions, so that predictable sets of benefits and costs are treated depending on whether a program is a dam, a transportation system, a training program, a government regulation, or whatever. What-ever the sources, it is clear that it is the analyst's responsibility to look, beyond the obvious gains and losses, to factors that are more indirectly or distantly implicated but are nevertheless important. Gramlich (1990) mentions such factors as pollution, health, safety, waste of time, secondary market impacts, and impacts on marital and family ties (Mohr, 1995).

At the same time, on the limiting side, analysts are cautioned to be sensitive to the possibility of duplication, or double counting, by including the same cost or benefit more than once in different guises. Double counting of benefits or costs must be avoided (Mohr, 1995).

The valuation of benefits and costs should reflect preferences revealed by choices which have been made. The most challenging part of Cost-Benefit Analysis is finding past choices which reveal the tradeoffs and equivalencies in preferences. For example, the valuation of the benefit of cleaner air could be established by finding how much less people paid for housing in more polluted areas, which, otherwise, was identical in characteristics and location to housing in less polluted areas.

It is sometimes necessary in Cost-Benefit Analysis to evaluate the benefit of saving human lives. There is considerable antipathy in the general public to the idea of placing a money value on human life. Economists recognize that it is impossible to fund every project which promises to save a human life, and that some rational basis is needed to select which projects are approved and which are turned down. The controversy is defused, when it is recognized that the benefit of such projects is in reducing the risk of death. This computation is equivalent to placing an economic value on the expected number of lives saved.

In fact it is recognized that not all impacts can successfully have a monetary value placed on them. Analysts are urged to go as far as possible in that direction and, rather than ignore not monetized impacts, simply present them on the side, as it were, in their original scales (Gramlich, 1990). Another weighting that may become important for the Cost-Benefit Analysis is the weighting of groups for importance in connection with the impacts of a particular policy. This kind of weighting is done subjectively by the analyst, often in consultation with others who have a concern regarding the policy. In this sort of case, the evaluator usually applies techniques such as sensitive analysis, that provide decision makers with a series of results instead of only one, each depending on the use of different set of weights (Mohr, 1995).

The function of assessing impact is central to Cost-Benefit Analysis. It is performed in highly complex ways and demands a great deal of expertise and creativity. The predominant approach involves a) the acquisition of at least some minimally essential amount of real world data on prices and quantities (including information about

goods, wages, interest rates, etc) and b) economic theory as a basis of extrapolating and imputing from there (demand and supply curves, marginal cost curves, etc) (Mohr, 1995).

The impacts of a project are defined for a particular study area, be it a city, region, state, nation or the world. The nature of the study area is usually specified by the organization sponsoring the analysis. Many effects of a project may «net out» over one study area but not over a smaller one. The specifications of the study area may be arbitrary but they may significantly affect the conclusions of the analysis.

#### 4.2.1.2 *Cost Effectiveness Analysis*

Cost Effectiveness Analysis is a feasible alternative to Cost-Benefit Analysis when benefits cannot be calibrated in monetary units. It permits programs with similar goals to be compared in terms of their relative efficiency and can also be used to analyze the relative efficiency of variations of a program (Rossi and Freeman, 1993).

But, because the benefits are not converted to a common denominator, we cannot ascertain the worth of merit of a given intervention in monetary terms from such an analysis. Likewise, we cannot determine which of two or more programs in different areas produce better returns. In this analysis, efficiency is judged by comparing costs for units of outcome (Rossi and Freeman, 1993).

#### 4.2.2 *Impact Analysis*

Impact assessments are undertaken to determine whether a program has its intended effects. Such assessments may be made at any stage of the program, from pre implementation policy making through planning, design and implementation (Rossi and Freeman, 1993).

Impact assessments may make use of qualitative or quantitative data. Although qualitative data are important for certain evaluative purposes, precise assessment of impact requires carefully collected quantitative data (Rossi & Freeman, 1993).

The main deficiency of the Impact Analysis design is that ordinarily it cannot disentangle the effects of extraneous factors from the effects of the intervention. Consequently, estimates of the intervention's net effects are dubious at best (Rossi and Freeman, 1993).

An additional complication is that, when programs have been in place for a period of time, «before» measures normally can be ga-

thered only by asking participating targets to reconstruct retrospectively what they were like before the intervention. In such studies the unreliability of recall can be a serious design effect.

As far as the Time Series Analysis is concerned, there are some limitations in its application. Perhaps the most serious limitation of many Time Series designs is the large number of pre-intervention observations needed in order to model pre-intervention trends accurately. Indeed, a Time Series Analysis can be performed only if extensive before-enactment and after-enactment observations on outcome measures exist. Of course, for many ongoing interventions, such long term measures do not exist. For this reason Time Series Analysis is usually restricted to outcome concerns for which governmental or other groups routinely collect and publish statistics (Rossi and Freeman, 1993).

#### 4.2.3 *Multicriteria Analysis*

Multicriteria Analysis compares the effectiveness and efficiency of several projects or interventions. It takes into account and synthesizes several effects, quantitatively and/or qualitatively. Multicriteria Analysis can be used to take into account any differences in the points of view of the partners. It is generally used *ex ante*. Its flexibility makes it suitable for use in many situations (European Commission, 1999).

The methodologies, presented previously, are tabulated in a critical assessment in Table 1.

## 5. Proposal of methodological approach in program evaluation

In the previous chapters we have proclaimed that each evaluation methodology has as many advantages as weaknesses. At this point, based on the above analysis, we propose two methodological approaches in program evaluation. In particular, we propose either the parallel application of several methodologies in an assessed program, or the mixing of existing pure methodologies to create a new approach. We appreciate that these two approaches can help the evaluator to overcome the weaknesses that may arise during the implementation of a single pure methodology.

**Table 1**  
**Comparative Assessment of Evaluation Methodologies**

Methodology	Origins	Qualitative or Quantitative Approach	Efficiency or Equity Criterion	Advantages & Disadvantages	When to use
Cost-Benefit Analysis	Jules Dupuit (A French Engineer) on 1848	Quantitative	Efficiency estimation	A: Of general acceptance D: Difficult to monetize qualitative variables	In standard types of projects
Cost Effectiveness Analysis	Originates from certain types of evaluations, notably in the domains of health and road safety	Quantitative	Efficiency estimation	A: Can serve to compare programs D: Compared programs should be of the same type	Specially in ex post evaluation
Impact Analysis	First used to define strategies or the re-organization of firms	Qualitative & Quantitative	Efficiency estimation	A: Allows a before-after comparison D: Failing of discerning the extraneous effects	In all phases of a program
Planning Balance Sheet	Lichfield, 1975	Qualitative	Equity Principles	A: Provides a general view D: Non focused	For descriptive analysis and when flexibility is desired
Goal Achievement Matrix	Hill, 1966	Quantitative	Efficiency estimation	A: Analytic assessment of goals D: Sometimes not precise	When decision makers are participants
Multicriteria Analysis	Nijkamp, 1975 (Concordance Analysis) Holmes, 1972 (Ordinal Ranking Methods)	Qualitative & Quantitative	Efficiency estimation & Equity Principles	A: Balance between conflict interests and possible ranking of needs B: May incorporate subjective elements	When dealing with more interests

Source: Own editing.

## Triangulation

Triangulation refers to the use of different data collection techniques and different research strategies to study the same program. According to Patton (1995), the term Triangulation is derived from the geometric shape triangle. The triangle is the strongest of all the geometric shapes and triangulated evaluation designs are aiming at increasing the strength of any evaluation. It is in data analysis that this strategy of triangulation really pays off, but in this paper we will discuss using multiple methods to study a program.

Triangulation is a powerful solution to the problem of relying too much on any single method and thereby undermining the validity and credibility of findings because of the weaknesses of any single method. Triangulation is the recognition that the evaluator needs to be open to more than one way of looking at the program (Patton, 1995).

Triangulation may be an ideal approach for evaluating a program. However, it is also very expensive. Most evaluation research involves quite limited budgets, short time frames and political constraints. In the reality of limited resources, attempts at triangulation may mean a series of poorly implemented methods rather than one approach well executed.

Besides, in the case of operational programs, Economou (1997) reports that a «synergy of the evaluations» is necessary. Indeed, since the evaluation of a specific program cannot include all related programs, there are obvious limits to the autonomy of every single evaluation. This leads to the conclusion that the evaluations of the different programs, even if they could follow a distinct time-table from an internal point of view, have to obey an overall co-ordination that will create the necessary inputs and outputs from/to the complementary interventions (Economou, 1997).

## Mixing analysis approaches

Triangulation is one way of increasing methodological power. While a second one is to borrow and combine parts from pure methodologies, thus creating mixed methodological approaches (Patton, 1995).

We believe that there are strengths and virtues in the ideal of pure implementation of each approach. But there are also some important benefits to be gained by mixing methods and approaches.



For example when dealing with a program in which some of the variables to be examined are unclear. It is possible to design an experimental design for collecting some qualitative open-ended data from the program participants.

The analyst may superimpose quantitative scales and dimensions onto qualitative data. Thus in the data analysis phase of a project the evaluator may decide to convert qualitative descriptions into quantitative scales that can be statistically manipulated.

A variety of mixes may then be done, such as mixes of data type, inquiry mode and analysis methods. In order to make the choices available more clearly, we list below some possible combinations for the creation of a flexible evaluation design (Patton, 1995):

1. Experimental Design – Qualitative Data Collection – Content Analysis
2. Experimental Design – Qualitative Data Collection – Statistical Analysis
3. Naturalistic Inquiry – Qualitative Data Collection – Statistical Analysis
4. Naturalistic Inquiry – Quantitative Measurement – Statistical Analysis

It should be mentioned that it is not possible to combine all the methodological approaches in a program. Certain designs pose constraints that exclude other possibilities. For instance, as mentioned above, it is possible to convert detailed qualitative descriptions into quantitative scales for the purposes of statistical analysis. However, it is not possible to work the opposite way, to convert purely quantitative methods into detailed, qualitative descriptions (Patton, 1995).

The choice of the appropriate methodological combination for the evaluation of a program depends on several factors, such as the purpose of the evaluation, what the stakeholders want to know, the funds available and the skills of the evaluator. It is certain that different methods produce quite different information. The challenge for the evaluator is to find out which information is most needed and most useful in a given situation, and then to employ those methods best suited to producing the needed information (Patton, 1995).

## 6. Conclusions

In this paper we have attempted to gather and summarize the basic methodologies in program evaluation. We have presented the major methodological approaches. After impressing on the main characteristics, advantages and weaknesses of the approaches and methodologies we proceeded to a comparative assessment. Eventually, we proposed Triangulation and the appropriate mixing of pure methods as two new strong methodological approaches.

From the above analysis we can conclude the following:

The impact of full coverage socio-economic programs is difficult to assess with confidence because the nature of such programs precludes the use of comparison groups. Full coverage programs generally are evaluated by using reflexive controls to compare preprogram and post-program outcome measurements. Designs range from simple before/after test evaluations, in which there is only one measurement before and one after program implementation, to time series evaluations, in which there are multiple measurements before and after the intervention is in place. Time series designs are much more powerful than simple before/after designs in estimating net effects. In evaluations with only two measurements, it is almost impossible to differentiate net from gross effects.

Cost Effectiveness analysis judges the efficiency of the intervention by quantitative comparison with one or more other interventions. It favors an objective considered to be a priority. It is particularly well suited to ex post evaluation.

Cost Benefit analysis judges the efficiency of the intervention by taking into account all its effects and by making a quantitative synthesis. Cost Benefit analysis can judge an intervention in the absolute, without any comparative reference to other intervention. It is generally used ex ante, but all the conditions required for its efficiency application are rarely met.

Qualitative evaluation data may be presented alone or in combination with quantitative data. Recent developments in the evaluation profession have led to an increase in the use of multiple methods including combinations of qualitative and quantitative data.

We have concluded that mixing parts of different approaches is a practical mandate to gather the most relevant possible information to inform decision makers and stakeholders. In practice, it is altogether possible and often desirable to combine approaches.

## References

- Economou D., *The Intermediate Evaluation of the Second Regional Operational Program of Thessaly*, 1997.
- European Commission, *Evaluating Socioeconomic Programs-Principal Evaluation Techniques and Tools* vol. 3, Office of Official Publications of the European Communities, Luxembourg, 1999.
- Gramlich E.M., *A Guide to Benefit-Cost Analysis* (2nd ed.), NJ: Prentice Hall, Englewood Cliffs, 1990.
- Hill, «A Goals Achievement Matrix for Evaluating Alternative Plans», *Journal of the American Institute of Planners*, 34, no 1, 1968, p.p. 19-28.
- Holmes J. C. (1972), «An Ordinal Method of Evaluation», *Urban Studies*, 9, no 2, p.p. 179-191.
- KEPE, *Reports 28, Assessment Study of the Regional Corporate Programs and the European Structural Fund 1994-1999*, Athens, 1997 (In Greek).
- Lagos D. and Lianos D., «The Scientific Approach of Greek Regional Planners Association for the Evaluation of Developmental Programs», in proceedings Evaluation of Operational Programs-Means-Methods-Techniques, *Greek Regional Planners Association*, Athens, 2000.
- Levin H.M., «Cost-Effective Analysis in Evaluation Research», in Gutten-tag M. and Struening E. L. (eds.), *Handbook of Evaluation Research*, vol. 2, Sage, Beverly Hills, 1975.
- Lichfield N. et al., *Evaluation in the Planning Process*, Pergamon Press, Oxford, 1975.
- Mohr L.B., *Impact Analysis for Program Evaluation* (2nd ed.), Sage, California; London; New Delhi, 1995.
- Nijkamp P., «A Multicriteria Method of Project Evaluation», *Papers of the Regional Science Association*, 45, 1975, p.p. 87-114.
- Patton M.Q., *Utilizatio – Focused Evaluation*, Sage, London, 1986.
- Patton M.Q., *How to Use Qualitative Methods in Evaluation*, Sage, Newbury Park; London; New Delhi, 1987.
- Rossi P.H. and Freeman H.E., *Evaluation – A Systematic Approach*, Sage, Newbury Park; London; New Delhi, 1993.

## Internet sites

[http://www.sjsu.edu/faculty/watkins/Cost-Benefit Analysis.htm](http://www.sjsu.edu/faculty/watkins/Cost-Benefit%20Analysis.htm)  
[http://www.mindtools.com/pages/article/newTED\\_08.htm](http://www.mindtools.com/pages/article/newTED_08.htm)



## The peculiar economics of doping in sports

*Manfred Tietzel\**  
*Christian Müller\*\**

«My compliments to my team and my thanks to the doctors.  
They achieved something superhuman.»

*Berti Vogts*, former coach  
of the German international soccer team

### *Abstract*

*The problem of doping in sports consists of a social dilemma situation in which, in equilibrium, each athlete consumes doping means although each one would be better off if everybody gave up taking performance-stimulating drugs. Based on a rational-choice analysis, a proposal which focuses on the «demand side» of the problem is made how this dilemma may be overcome. An abstract, non-enumerative definition of the doping offense is put forward which classifies all substances enhancing the athletes' performance and damaging their health as doping means. In a multilateral «fairness compact» among all participants involved, sports associations and promoters would have to commit themselves on a life-long suspension of all athletes tested positively. Sponsors would have to provide the financial basis for an independent monitoring and enforcement agency. It is argued that this proposal not only drains the doping problem in an incentive-compatible way but also avoids the inefficiencies and inhuman consequences of actual doping practices.*

---

\* Professor of Economics, Department of Economics, Gerhard Mercator University Duisburg, D-47048 Duisburg/Germany, E-mail: m.tietzel@uniduisburg.de

\*\* Lecturer in Economics Duisburg - Essen University.

## 1. Introduction

For several years the media interest in many kinds of high-performance sports seems to have focused as much on the use of drugs as on athletes' records. Positive doping tests and the sudden deaths of a number of athletes have fuelled further rumours. A popular and seemingly obvious explanation for the doping problem is that the athletes' morality and the long standing standards of fair play have gradually eroded with time. In this paper we shall try to show, by means of a rational-choice approach, that there is a different, deeper and, in our opinion, more adequate explanation of drug abuse in sports. Policy advice derived from this explanation differs from recommendations hitherto put forward.

## 2. Are there any good reasons to dope?

From a rational-choice perspective, the participants in any major sporting event are captured in a prisoners' dilemma, where individual rational behaviour leads to a jointly destructive outcome (see, for instance, Breivik, 1987). On the one hand, literally no one could be interested in using drugs as a means of doping in modern, top-level sports. Spectators wish to enjoy a fair competition; sponsors are interested in avoiding damage to their images; and those who have the least to gain from doping may be the athletes themselves who not only bear a high economic and, presumably, moral cost of doping, but also incur health risks by using performance-enhancing drugs. An enforced control on doping is in everybody's encompassing self-interest. Since each individual party can gain from this, fair play in sports is Pareto-superior to a situation of general drug use.

For any top-class athlete, however, it is individually rational not to comply with a doping ban. Suppose firstly a fictitious situation in which there are no doping regulations at all. What would a sportsman choose to do if he expected all competitors not to be doped? From his point of view, it would be rational to consume performance-enhancing substances in order to increase his particular chance of winning a tournament or, perhaps, of setting a world record. And how, on the other hand, would he act if he knew that his competitors used drugs? In this case, he would have another good reason to

use drugs himself. Since, dispensing with doping would be harmful to himself in that he would have to pay for his unilateral self-restriction by a highly probable defeat. In other words, if all others dope, the «nice guy will finish last» (Gardner, 1974). Therefore, taking one's daily dose of drugs is definitely dominant.

Since each and every athlete calculates in the same way, in equilibrium all participants find themselves in the collectively least desired all-doping situation. Given these circumstances, individual rationality implies group irrationality.

The chemical «arms race» for victory, yet, is socially inefficient. Only one of many competitors can win an event. The doping costs of each other athlete are wasted. In a situation in which drug use spreads to all athletes, a collective «disarmament» would be beneficial to all. Even the winner would gain since it is plausible to assume that a general «armistice» would not change the relative positions of the participants in a tournament. The athletes face the problem, however, of securing the reciprocity of fair play in order to be able to look after their encompassing self-interests, without the risk of ending up as a sucker. Just as Ulysses once let himself be bound to the mast of his ship in order to withstand the sirens' singing, athletes are in need of an external binding mechanism in order to be able to protect themselves from becoming objects of their own narrow interests. Therefore, an enforced doping ban does not bind the athletes to some arbitrarily chosen and externally imposed set of rules, but to their own encompassing interests.

In modern societies, situations abound in which only externally enforced restrictions on individual freedom can help a person to act in accordance with his own encompassing interest. All kinds of justifications of a government's intervention into its citizens' liberty in one way or other depend on a «paradox of being governed» (Buchanan, 1975). As Jean-Jacques Rousseau (1913: 15) put it 200 years ago, men must «be forced to be free.» A century earlier, Thomas Hobbes maintained, in a similar vein, that only «that great Leviathan» – a common power formed by, but external to a given state's citizens – can free men from their positional «warre of every man against every man» in anarchy, which renders the life of man «solitary, poore, nasty, brutish, and short» (Hobbes, 1914: 65).

External coercion can help to avoid an anarchical competition fuelled by the participants' partial interests. In the American Hockey League, for example, players deliberately chose not to wear helmets, even though most of them explicitly claimed to favour a

mandatory rule to do so (see Schelling, 1978). The reason for this was that many players believed that wearing helmets would reduce their efficiency and put them at a disadvantage, since they would be handicapped by their head protection. Thus, every player valued avoidance of an inevitable reduction in competitive strength more highly than his health. In this example, hockey players are caught in a social trap: their strong taste for victory causes a jointly harmful outcome. If wearing helmets was mandatory, all players could achieve a joint cooperative situation in which no one would have an advantage over the others, and at the same time everyone would incur a smaller risk of being injured.

### 3. Under which conditions will athletes use drugs?

External enforcement systems of doping regulations are already in existence. Why do the athletes dope nevertheless? One of the reasons may well be the fact that laws are never perfectly enforced. Provision of information about the actions of individuals is costly and must respect their privacy. Not every single violation of a law will, therefore, be sanctioned (see Stigler, 1970). The effective amount of sanctioning depends both on the punishment laid down in the law and on the probability of law enforcement.

A rational law-breaker, who maximizes the expected value of his net utility, will act illegally as long as the incremental utility of his action exceeds the expected marginal disutility imposed by sanctioning. This disutility is the product of the individually perceived costs of being sanctioned times the probability of discovery and punishment for the offense. In the above example of the American Hockey League, the probability of being punished for not wearing a helmet would have been near one, as it can be easily observed whether a player is wearing a helmet or not. A relatively small absolute sanction would, therefore, have been sufficient to help the players escape their social dilemma.

The case of doping is a very different matter. On the one hand, drug abuse by athletes is extremely difficult to observe; the use of certain drugs cannot even be detected at all. Consequently, the probability of detecting a violation of anti-doping rules is small. As comparatively low expected costs of punishment are coupled with



high prizes for winners, even a high nominal sanction may not suffice to suppress doping entirely.

Interpersonal differences in the expected value of a given sanction may explain why seemingly not all top athletes have equal incentives to dope. Some of them are bound by sponsoring contracts stipulating their immediate dismissal in case of a positive doping test. If such an additional amount of sanctioning in a sponsoring contract is high enough, a rational athlete will refrain from doping.

#### 4. The incentive to dope: Some hypotheses

##### Which athletes consume performance-enhancing drugs?

As alert readers of newspapers, we probably have the *prima facie* impression that doping is much more common in certain types of sports such as cycling or track and field athletics than in others such as golf or soccer. This impression, however, may in fact be grossly misleading. There is good reason to believe that we shall never be able to verify this impression nor its contrary. It lies in the very essence of doping that it be done in strict secrecy and, as a consequence, reliable and comprehensive data will never be available.

For one thing, the absolute number of reported doping cases in some branches of sports has a meaning only in relation to the number of athletes active in these fields of sports. Secondly, the incentives, as well as the particular characteristics of the various types of sports can differ to a large extent. For example, the difference in the incomes of the heavy weight champion in professional boxing and the biathlon world champion is certainly enormous. In addition to this, both, the number of doping tests per year and per athlete and the rules of sanctioning, differ from country to country and from sport to sport. Finally, the effect of drugs on performance is enormous in weight-lifting and probably close to zero in golf. It is for these reasons that any number of positive doping tests that should become known, do not indicate an effective doping intensity in the respective branches of sports. Since the available data, for these reasons, must be taken to be strongly and systematically biased, it would be inadequate to argue on such a basis. Under these informational conditions, all we can do is try to make some 'infor-

med guesses' about the plausible out-comes of different incentive structures.

As a first guess one might presume that, other things being equal, the incentive to dope is stronger in individual sports than in team sports. The reason for this seems to be obvious: whereas the benefits of an enhanced performance accrue exclusively to a doped heavyweight boxer, a doped football player's performance is beneficial for the whole team; the latter's doping would raise his contribution to a public good. Competition among teams for ranking high in their respective leagues causes only a moderate incentive for an individual team member to dope. As a member of a small group, his contribution to his team's success is certainly not insignificant nor is his resulting share in the increased team income. Normally, however, his performance is not decisive for the team's overall performance. Triumphs as well as defeats are always shared by all. Consequently, the individual incentive to smooth the team's way for victory by means of pharmaceutical products will be a negative function of group size: the bigger the team, the smaller the incentive to dope.

Alas, the moderate influence of team competition on an individual team member's decision to dope may well be amplified by competition among athletes to become or remain a member of a high ranking major league team. Since the number of high income jobs in major league teams is small and cannot be increased, competition for membership in a team must be expected to be extremely fierce. Contrary to intuition and casual observation, we should not expect team sports to be havens of fair playing guys (and girls for that matter).

Also, one might plausibly assume that the incentive to dope in individual sports (*ceteris paribus*) is more distinct whenever peak performances can be measured in absolute numbers. The actual world record holders, say, in long distance running or javelin throwing clearly rank higher than any comparable athlete in human history. Any current world record holder in such a branch of sports holds a worldwide natural monopoly, sometimes for quite an extended period of time as, for instance, did Bob Beamon who held the world record in long jumping for 23 years. This natural monopoly, if marketed appropriately, can create considerable rents.

In branches of sports in which results can only be recorded as a rank order of outcomes in a certain tournament, no such monopoly rents can be produced quite simply because a tournament winner

ranks first over a small number of competitors at a given place and time and not generally over all athletes in history. We have only a vague idea of who was the best golf player of all time or even of the last decade. Therefore, one should expect the ratio between the yearly income of a world record holder and that of the second ranking athlete to be higher than the income ratio of the athletes who won the most and those who won the second highest number of tournaments in a year. As a consequence, positional competition should be more pronounced and the incentive to dope more marked in the former than in the latter.

## 5. Which governments support doping?

Casual observation as well as a-priori reasoning suggest that the doping phenomenon is not only unequally frequent in different kinds of sports but also differs among states. Past and present communist states had and still have a reputation for systematically doping their athletes, which was supported by a purposeful pre-selection of talented youths. Recently sports officials of the former German Democratic Republic had to defend themselves in doping trials before the courts. These trials produced evidence that very young athletes were quite often doped against their will and, in many cases, even without their knowledge.

Again, reliable data of approximate completeness will never be available. As doping was implemented with governmental assent and support, everything was done to keep the real source of the international superiority of their athletes secret; with only a few exceptions of occasional circumstantial evidence, the facts will remain behind a thick veil of silence.

Having had very few, if any, social or economic performance characteristics comparable to those of Western democracies, being ahead of the rest of the world at least in the realm of sports was one of the rare and badly desired occasions to demonstrate what «socialist progress» means.

In order not to lose the self-initiated «competition between systems» due to an emigration drain, the citizens had to be kept within the national boundaries of the communist states by means of an unprecedented migration barrier, the Iron Curtain (see Tietzel and Weber, 1994). In order to win the competition between systems, at

least in the field of sports, top athletes primarily had to be tied closely to their socialist states of origin. On the one hand, they had to perform at international meetings; but, on the other hand, their incentives to flee from their team in order to have a chance at an international career had to be curbed. This was achieved by taking the sportsmen's relatives as hostages so to say who did not have the slightest chance to leave the country.

Therefore, in spite of outstanding performance, the income of a top Western athlete was unavailable to them. A home-made socialist incentive to dope was created by rewarding successful athletes with comparatively high incomes and many other privileges, for themselves and their families, such as preferential access to university education and housing. Given the comparatively low incomes of normal citizens in ordinary professions, the difference in income between a socialist top athlete and the average citizen was certainly extraordinarily large. In view of the relative size of this rent and owing to the fact that an athlete definitely was not sanctioned for doping, one should expect a high degree of consent among athletes to be manipulated for performance; in a way, they were in the positions of «big enchanted princes in small ponds» (see the cartoon in Frank, 1985: II). What was unachievable with hammer and sickle was made possible with the help of the syringe.

## 6. An erosion of morality?

According to a widely held opinion, fair play in sports was much more common in former times than it is nowadays. Given our analysis, this claim lacks any foundation. Through the ages, athletes have competed by any means for fame, glory and money. While today doped athletes, by manipulating the probability of victory, only harm their undoped competitors in an indirect way, athletes of former times even tried to reduce their rivals' competitiveness directly. In the ancient Greek agones boxers concealed metal weights in their hand belts or they fitted them with dangerous thorns. The historian Dio Cassius reported that the Roman emperor Caligula did not even refrain from poisoning the racehorses and the charioteers of those stables which were serious competitors to the «green» chariot racing team he favoured fanatically. The myth propagated by Pierre de Coubertin, the founder of the modern Olympic movement, that ta-

king part and not winning was important at the Olympics of antiquity, is certainly based on the 19th century's idealized view of antiquity, not however on historical facts.

A comparable historical misinterpretation seems to be the cause of the repeatedly expressed ad hoc explanation of doping cases as being the consequences of a widespread erosion of morality in sports. The observable steady increase of registered doping cases through the years can be explained in quite different terms. On the one hand, in many countries severe doping regulations were only introduced during the last decade; numerous kinds of medication formerly permitted have been banned since then. This alone may well have increased the number of doping offenses officially disclosed. On the other hand, the technical development of performance-enhancing substances has proceeded at an increased rate. While doping laboratories have been working hard to finally develop a reliable test for the use of the drug Erythropoietin (EPO), which still plays a prominent role in sports that require endurance, athletes allegedly have switched to a more efficient but also more dangerous substance. Given the rapid development of novel drugs, even the toughest of doping controllers will perhaps win a battle against drug abuse in sports; but they will never win the war.

Finally, the stakes within the big «game of sports» have risen significantly during the last few decades. At the beginning of the century, prominent sporting events, such as the Olympics or the Tour de France, were competitions among amateurs. These athletes were lured to take part, in the absence of high prizes and without any lucrative promotional offers, by the potential fame and possibly for entertainment, which, by the way, is the original etymological meaning of the word «sports». In such «low-cost situations» in which moral behaviour was relatively cheap, the incentive to cheat was not as strong as it is today.

The world of sports, however, has changed significantly since that time. The advent of modern media turned sporting events into goods of mass consumption. Whereas tournaments could formerly only be watched by spectators in a stadium, TV serves millions at zero or minimal variable costs. The decrease of per capita costs of an event's provision is no longer limited by the capacity of a stadium, but continues over the much larger numbers of sports fans at their TV sets.

Promoters of big sporting events fully exploit the commercial potential of sports. Spectators' interest in amateur sports has dimini-

shed, since access to high performance professional sports has become so easy and cheap. The admission of professionals to the Olympics and other international sporting events has certainly been motivated by the economic interests of their promoters. Under the pressure of attracting media interest, promoters compete with hitherto unexperienced rigor for the elite of international top-level sports.

The advertising industry discovered the promotional value of sports. A spectator, instead of buying a ticket, now «pays» in terms of simultaneous consumption of advertisements. Top ranking athletes are in a position to «lease» their natural monopoly to sponsors for commercial use. Advertising one's brand on an athlete's cap or shirt has a twofold advantage over other kinds of publicity. Whereas ordinary TV commercials attract only moderate attention, to say the least, advertisements on an athlete's clothing are naturally noticed jointly and with the same focus as this athlete's performance. In addition, the per capita costs of this type of ad decrease at a much higher rate than in other forms of advertisement because the amount of money the sponsor pays an athlete can be regarded as a fixed cost of advertisement. It is, therefore, hardly astonishing that the budgets of sports ads are constantly increasing and that competition among sponsors for top athletes is tough. In track and field athletics, most world records occur at major events in New York, London or Zurich because promoters and sponsors can, in some sense, «buy» a world record by offering a remuneration.

Sponsors and promoters are trapped in a social dilemma comparable to the athletes' doping dilemma. While the latter compete on streets and sports fields for records and titles, the former bid for the commitments of the prospective record-setters or for advertising contracts with the record-holders in a less visible way. In a positional competition barely less tough than that among the athletes, marketing agencies try to outbid each other with the result of ever increasing sponsoring budgets.

The mass commercialization of sports releases a «superstar effect» which, in turn, makes the reward structure in most sports highly nonlinear: the top-level athletes' rewards rise in great disproportion to talent and ability. As a consequence, the bulk of the earnings goes to a handful of participants. The right-hand tail of the income distribution among athletes is much thicker than the lower. For example, the top five money winners on the pro golf tour have annual stroke averages that are less than five percent lower than those of the fiftieth or sixtieth ranking player; however, they earn

four or five times more money. Thus, trifling differences in performance sufficient to define the rank order may cause substantial differences in incomes (see Rosen, 1983: 451). The American freestyle swimmer Shirley Babashoff, for instance, won four silver medals in the Montreal Olympics in 1976 although she was the favourite. For her events, she was beaten by East-German swimmers, whom she described in an interview as «huge, hairy and deep-voiced. ... I was not second, I have been defeated by men.» Today she works as a postwoman in Huntington Beach, Los Angeles, and she still grieves over the idea, that as a fourfold gold medal winner she could have expected a career comparable to Mark Spitz's (see Kalwa, 1998).

It is simply insufficient even to be ranked a very close second place. As the often cited slogan, attributed to Vince Lombardi, puts it: «Winning isn't everything, it is the only thing» (quoted in Breivik, 1987: 84). If a great deal is at stake the athletes tend to do anything to win, doping included. This has less to do with an erosion of the athletes' morality and more to do with an incentive system in professional sports which has changed radically. Within the complicated nexus of interests involving spectators, sponsors, promoters and managers, the price athletes have to pay for morality, these days, has simply risen to extremes.

## 7. An incentive compatible solution to the doping problem

### «Doping» defined

Are there safe ways out of these dilemmas? The difficulties in finding a solution begin with the very problem of stipulating the types of behaviour that are to be regarded as «doping.» The usual expedient, adopted for instance by the International Olympic Committee, is to set up a so-called «negative list» which specifies all the substances the use of which is to be regarded as a fact constituting «doping.» The use of drugs not included in the negative list is not liable to be punishable. This provides a strong incentive for athletes, coaches and sponsors to substitute the listed drugs for novel, unlisted drugs in order to avoid doping in the legal sense, yet not factually. The incentive to substitute illicit drugs for novel ones correlates with the degree of enforcement of the negative list. As parado-

xical as it may sound, a strict enforcement of a negative list will aggravate the doping problem.

In order to avoid evading strategies and their perverse side effects, it has been suggested that the athletes themselves be entrusted with defining what is to be regarded as doping (Bird and Wagner 1997). According to this proposal, any drug which enhances an athlete's performance so sweepingly that he is unwilling to overtly concede its consumption is to be considered a doping substance. The core of this proposal amounts to demanding that all athletes keep a drug diary in which they record any drugs taken. If, by means of comprehensive tests, it becomes evident that an athlete has taken a drug not enlisted in his diary he is found guilty of doping and will be punished. The premise upon which this proposal rests is that the term «doping» now refers not to the consumption of a drug on a negative list, but to drugs used secretly.

It seems highly questionable, though, that the evading reactions triggered by a negative list will be removed by means of drug diaries. The range of drugs, the overt use of which is not regarded as embarrassing by the athletes, will gradually expand. An athlete, who refrained from taking certain drugs up to some point of time and who, more and more, observes others frankly confessing the use of these drugs, will experience a growing competitive disadvantage. The decision to use the drug himself will become all the more likely the less the general public is informed about the effects of the drug and the less reason he has to feel embarrassed about using it himself; what so many others do cannot be completely disrespectable. Thus, the introduction of the drug diary will initiate a continuous expansion of the number of drugs used overtly. It will not lead to a removal of doping, but to its factual legalization.

Therefore, one should not be reluctant to define doping in an abstract way, even though the problem of operationalization will be unavoidable in the sense that cases will occur where it will be difficult to decide unambiguously, whether these represent the constituent facts of doping or not. No legislator would even consider the idea of defining «theft» by means of a negative list of the tools needed for burglary or allow the perpetrators to decide which tools are admissible for burglary and which are not.

It seems sensible and also operational to define «doping» as an athlete's consumption of drugs which increase his competitive performance and at the same time are damaging to his health. If the use of some drug were not performance-enhancing, no third party



would be placed at a disadvantage; no other athlete would feel compelled to take part in a chemical arms race. If the drug were not injurious to one's health, its use would hardly be distinguishable from one's daily intake of food. Given this definition of doping the burden of risk to be punished for taking novel drugs is completely with the individual athlete, whereas an exhaustive negative list does not impose such risk at all.

According to our definition, even those athletes whose performance was manipulated by some third party without their knowledge are to be regarded as doped. Examples that come to mind are the manipulation of an embryo's genes or, as was a widespread practice in the former German Democratic Republic, the administering of hormones to child athletes. These athletes, too, must be punished, by excluding them from sports competitions, in order to avoid other athletes being placed at a disadvantage and to obviate the incentive for these other athletes to keep up with them. An athlete doped without his knowledge can, of course, sue the manipulator for the damages caused by the assault.

## 8. A fairness compact in sports

In addition to defining «doping» in a clear-cut way, it has to be established who is to be punished, to what extent and in which way. The regulations against doping, be that those in existence or those called for by officials, relate mainly to the «dealers» i.e. coaches or doctors who supply and provide the drugs. According to the German Pharmaceutics Act of 1998, for instance, it is explicitly forbidden to «offer for sale, prescribe or administer to others pharmaceuticals for the use of doping in sports.» Violations can be punished by a fine or imprisonment of up to three years. Also, just recently, the Australian Olympic Committee did not hesitate to call for a lifelong period of detention for such an offense.

Doped athletes are usually treated with extraordinary leniency; sports associations consider suspensions of a few months to be an adequate punishment. Even the Australian Olympic Committee which calls for punishing dealers in a draconian way contents itself with a two years punishment for athletes tested positively.

Imprisonment, be that of athletes or suppliers, for violations of doping regulations is as inadequate a punishment as it is inefficient.

To begin with, a person absolutely uninvolved, the general taxpayer, would have to bear the direct cost of executing the sentence. In addition, punishing the act of doping by imprisonment is in harsh disproportion to the nature of the offense. Doping an athlete is something quite different from an offense such as killing him, all the more so in the case of an athlete who takes drugs voluntarily. Finally, whether or not the legislator is the most appropriate regulator in this area can be reasonably questioned; there is good reason to believe that sports associations, sponsors and promoters of sports events are the more natural and more efficient regulators.

It is conceivable that there is a solution to the doping problem which is incentive-compatible, efficient and which preserves the autonomy of sports. This solution could consist in a fairness compact among sports associations, promoters and sponsors that ideally should comprise all kinds of sports and should be negotiated on a supranational basis. The main emphasis of that compact should be the regulation of the sanctions on athletes, who are found guilty of doping, and the enforcement of such sanctions. The compact should be designed in such a way that it is self-enforcing in the sense that it lies in the self-interest of the contracting parties to comply with the rules laid down. In addition, a unanimous agreement on the rules of the compact should be conceivable, as well as reasonable, in the sense that everyone who is party to it would be better off with it than without it. Contrary to actual practice and alternative proposals, the fairness compact we have in mind aims at the demand side for drugs; the suppliers, who are nowadays the only ones who face a serious threat of sanctions, go unpunished under such a system of rules with the obvious exception of the ordinary legislation concerning assault.

First of all, sports associations and promoters should agree on sanctioning athletes convicted of doping much more seriously than now. Athletes who have tested positively for doping ought to be excluded irrevocably from all forms of sports and from every competition for a period of time as long as their expected sporting career. A lifetime exclusion would be unambiguous and render unnecessary complicated calculations and predictions. Though lifetime exclusion for athletes may seem an unusual sanction and unappropriate given the nature of the offense, it is not without precedent; in the fifth century B.C. Greek athletes, who fatally injured their opponents in boxing, were banned for the rest of their lives from the Panathenian games (see Connolly and Dodge, 1998: 83).

Such a provision, if perfectly enforced, would reduce a professional athlete's expected lifetime income to that which is achievable in the next best job he is able to perform. The rent on being a gifted professional athlete would be reduced to zero, and the athlete alone, who makes the decision to dope and possibly profits from it, would incur the cost of being sanctioned which is measurable in terms of a reduction in his expected lifetime income.

An exclusion from sports competitions for the expected duration of an individual's sporting career, rather than imprisonment, is perfectly incentive-compatible insofar as —due to the superstar-effect— the expected cost of being sanctioned increases more than proportionally compared to the incentive to dope. The expected cost of being sanctioned for doping is higher, the higher an athlete ranks in his branch of sports and the better, for this reason, his possibilities are to obtain a high income as a professional athlete.

Insofar as it can be assumed that the rank order of athletes is the same with and without doping, this part of the fairness compact makes an improvement feasible, to which the athletes' unanimous consent can, with good reason, be assumed.

In a similar vein, sponsors, too, could make an improvement if they, on their part, would bind themselves to sustained efforts to increase the probability of detecting doping offenses. In order to create a reputation of supporting fair and clean athletes for themselves, they would be well-advised to spend a fair share of their sponsoring money on financing independently conducted doping controls and on the development of new and improved doping tests.

Information on how much of a single sponsor's advertising budget is contributed to the support of doping testing and detection, should be made public in order to build an individual reputation for fairness. The general public will regard these individual contributions as a pledge by means of which a sponsor makes his reputation dependent on the behaviour of the sponsored athletes. If any of the sponsored athletes were tested positively for doping, the pledge would be lost. Therefore, sponsors would have an incentive to urge their athletes to compete fairly when they negotiate sponsoring contracts.

The suggested committing compact is credible only if doping tests and legal procedures are implemented by independent institutions. If controls or enforcement activities were carried out by sponsors, promoters or sport associations themselves, top athletes might put pressure on them by threatening to refrain from participating in

a sports meeting unless left untested. Even in the case of an athlete found guilty of doping, sponsors as well as promoters would have a strong incentive to suppress information about the matter in order to retain their respective reputations and pledges. In the existing system this has already happened; in 1996, for instance, in the Olympics held in Atlanta, the IOC made efforts to suppress information on eleven positive tests.

## 9. Conclusion

The doping problem in sports is solvable. It consists of a prisoners' dilemma-type situation in which, in equilibrium, each individual sportsman acting rationally consumes performance-enhancing drugs although each one would be better off if everybody refrained from taking them. The cause of this dilemma is the athletes' commitment problem which turns their individual choices of fair play into a self-damaging choice, as long as it cannot be secured that all other athletes behave correspondingly.

In this paper we have launched a proposal which allows all those affected to escape from this dilemma. Other than «negative lists» scheduled by the IOC or other sports associations, our proposal is based on an abstract, non-enumerative definition of the doping offense which considers all performance-enhancing and health-damaging substances as doping means.

In contrast to other proposals and legal regulations which focus on the «supply side» of the problem we suggest, on the basis of a rational-choice analysis, a system of incentives that aims at the «demanders» for drugs in sports. All participants –sports organizations, promoters and sponsors– could be better off in consenting to a comprehensive fairness compact. In such a compact sports associations and promoters would have to commit themselves to an extended suspension of all positively tested athletes. Sponsors would have to provide the financing of an independent monitoring and enforcement agency.

We have shown that this proposal not only drains the doping problem in an incentive compatible way, but also avoids the inefficiencies and inhuman consequences of actual practice. Whether the contents of a top athlete's baggage will continue to attract more media interest than his obtained results, which are often stunning wi-

thout doping, depends on how the doping problem will finally be settled.

## References

- Alchian, A.A. and H. Demsetz (1972), «Production, Information Costs and Economic Organizations,» *American Economic Review*, 62, 777-95.
- Bird, E.G. and G.G. Wagner (1997), «Sport as a Common Property Resource. A Solution to the Dilemmas of Doping,» *Journal of Conflict Resolution*, 41, 749-66.
- Breivik, G. (1987), «The Doping Dilemma. Some Game Theoretical and Philosophical Considerations,» *Sportwissenschaft*, 17, 83-94.
- Buchanan, J.M. (1975), *The Limits of Liberty. Between Anarchy and Leviathan*, Chicago: University of Chicago Press.
- Connolly, P. and H. Dodge (1998), *The Ancient City*, Oxford: Oxford University Press.
- Frank, R.H. (1985), *Choosing the Right Pond. Human Behaviour and The Quest for Status*, New York - Oxford: Oxford University Press.
- Gardner, P. (1974), *Nice Guys Finish Last. Sport and American Life*, London: Allen Lane.
- Hobbes, T. (1914 Reprint 1970), *Leviathan*, London - New York: Everyman's Library, 1914.
- Kalwa, J. (1998), «Die betrogene Briefträgerin will das falsche Silber verscherneln,» *Frankfurter Allgemeine Zeitung*, No. 286, 9/12/1998, 40.
- Rosen, S. (1983), «The Economics of Superstars,» *The American Scholar*, 52, 449-60.
- Rousseau, J.J. (1913 Reprint 1968), *The Social Contract and Discourses*, translated with Introduction by G.D.H. Cole, London - New York: Everyman's Library.
- Schelling, T.C. (1978), «Hockey Helmets, Daylight Saving, and Other Binary Choices.» In Schelling, T.C., *Micromotives and Macrobehavior*, New York - London: Norton, 211-43.
- Stigler, G.J. (1970), «The Optimum Enforcement of Laws,» *Journal of Political Economy*, 78, 526-36.
- Tietzel, M. and M. Weber (1994), «The Economics of the Iron Curtain and the Berlin Wall,» *Rationality and Society*, 6, 58-78.





# Testing the existence of an environmental Kuznets Curve for sulfur using panel data models

*George Emm. Halkos\**

## *Abstract*

*This paper uses a dynamic panel data for 73 (23 OECD and 50 non-OECD) countries for the time period 1960-1990 in order to estimate the relationship between Gross Domestic Product (GDP) and pollution in the form of sulphur emissions. For the purpose of performing our empirical estimate, fixed and random effects are used. This analysis shows significant differences between the most industrialized countries and the rest of the countries considered. This implies that a uniform policy to control pollution is not adequate. It is necessary to take into consideration the specific economic situation as well as the structure of the industrial and the business sectors in each country. Finally, in terms of policy implications, the study presents the main abatement options for sulphur reduction.*

## 1. Introduction

The generation of electricity from conventional power stations is associated with a number of environmental problems. For example,

---

\* Assistant Professor in Econometrics-Statistics, Department of Economics, University of Thessaly.

generation using coal causes significant air pollution due to emissions of sulfur oxides, carbon dioxide, nitrogen oxides and particulates. In the UK a 2000 MW coal fired station operating at 60% load factor burns about 4.4 million tons of coal per year and each year emission into the atmosphere is about 10 million tons of carbon dioxide, 130,000 tons of sulfur dioxide, 40,000 tons of nitrogen oxides and between 4,000 and 40,000 tons of particulate matter depending on how well the stack emissions are cleared before they are released (Highton and Webb, 1980). Particular concern has been expressed about the emissions of sulfur dioxide because the use of tall stacks to disperse emissions can lead to problems of transnational pollution. Approximately 1 ton of sulfur burned produces 2 tons of sulfur dioxide (SO<sub>2</sub>) and sulphur is present, in varying quantities, in both oil and coal.

Kuznets (1965, 1966) showed that during various economic development stages income disparities first rise and then begin to fall. The environmental Kuznets curve (hereafter EKC) hypothesis proposes that there is an inverted U-shape relation between environmental degradation and per-capita income. Environmental damage seems to be lower in the most developed countries compared to many middle-income countries and higher in many middle-income countries compared to less developed countries. It is worth mentioning that an alternative form of the EKC hypothesis suggests that environmental degradation as a function of income is not a stable relationship but may depend on the level of income. This is because in this alternative form, there may exist one relationship for poor and another for rich countries. On the aggregate this would give an inverted U-like curve.

Cropper and Griffiths (1994) and Selten and Song (1994) conclude that the majority of countries in their analyses are below their estimated peak levels for air pollutants and thus economic growth may not reduce air pollution or deforestation. This implies that estimating the left part of EKC is easier than estimating the right hand part.

A number of authors have estimated econometrically the EKC using OLS analysis. The use of OLS is not likely to yield accurate estimates of the peak levels. The EKC estimates for any dependent variable (e.g. SO<sub>2</sub>, NO<sub>x</sub>, deforestation, etc.) peak at income levels which are around the world's mean income per capita. Income as expected is not normally distributed but skewed (with a lot of countries below mean income per capita). Arrow *et al.* (1995), Ekins (1997) and Ansuategi *et al.* (1998) provide a number of reviews



and critiques of the EKC studies. Stern et al. (1996) identified a number of problems with some of the main EKC estimators and their interpretation. They mention among other econometric problems, the mean-median income problems, the interpretation of particular EKCs in isolation from other environmental problems, the assumption of unidirectional causality from growth to environmental quality and the reversibility of environmental change and asymptotic behavior. Stern (1998) reviews these problems in details and shows where progress has been made in empirical studies.

In addition to the fact that the use of OLS is not an appropriate technique in modelling the EKC, most of the empirical studies do not present diagnostic statistics of the regression residuals. Due to this reason we cannot be certain that the peak levels provided –and the policy implications suggested– are accurate. Halkos and Tsionas, (2001) using cross-sectional data, obtained the following results:

$$\text{Deforestation} = -2.344 + 1.298 \log \text{GNP} - 0.243 [\frac{1}{2}(\log \text{GNP})^2]$$

$$(0.7824) \quad (1.595) \quad (2.189)$$

Harvey test for heteroskedasticity  $\chi^2(2) = 9.213$ ,

RESET test for misspecification  $F(3, 55) = 3.03$

BP test for heteroskedasticity  $\chi^2(2) = 1.427$

Jarque-Bera Test for Normality:  $\chi^2(1) = 4.67$

Where t-ratios are presented in parentheses. These results indicate the existence of an EKC. However, the diagnostics imply the specification is totally unreliable as we see heteroskedasticity, misspecification and non-normality problems.

In this paper, we examine the concept of an environmental Kuznets curve in a critical way with an eye towards proposing policies compatible with sustainable development. A dynamic panel data for 73 countries, for the time period 1960-1990, is used in order to estimate the relationship between Gross Domestic Product (GDP) and pollution in the form of sulphur emissions. Our empirical estimation is performed using fixed and random effects. To control for non-observable specific effects Two Stage Least Squares (2SLS) is applied.

The paper is organised as follows: Section 2 discusses the existing theoretical and empirical work. Section 3 presents the data used while section 4 discusses the econometric models. The empirical evidence is presented in section 5. Section 6 describes briefly the abatement options for sulfur emissions reduction. The final section concludes the paper.

## 2. Previous work

The empirical analysis of the EKC has focused on whether a given index of environmental degradation shows an inverted-U relationship when it is related with income per capita. A number of possible explanations exists for the inverse U-shape relationship. Natural progression of economic development goes from clean agricultural to polluting industrial and to clean service economies. The argument here is that «scale effect», in the sense that more output results in more adverse effects for the environment, is (at least partly) offset by the «composition effect» due to the changes in the structure of the economy as well as the «technology effect» due to possible changes in the production methods. The improvement in environmental quality may be the result of the change in the technological mode of production (de Bruyn, 1997; Han and Chatterjee, 1997) or of the exportation of «dirty industry» to less developed or developing countries (Rock, 1996; Suri and Chapman, 1998; Heerink et al., 2001).

In the formalization of the transition to the low-pollution state there is a group of authors that provide significant analyses of the role of preferences and regulation on the emissions profile of polluters (Lopez, 1994; McConnell, 1997; Stokey, 1998). Dinda et al. (2000) claim that technological improvements, structural economic change and transition, increase in spending on environmental R & D accompanied with increasing per capital income are important in determining the nature of the relationship between economic growth and environmental quality.

Another explanation is that since air pollution is considered an externality, internalization of this externality requires relatively advanced institutions for collective decision making. This can be achieved only in developed economies. A better institutional set up in the form of credible property rights, regulations and good governance may create public awareness against environmental degradation (Dinda et al., 2000). Jones and Manuelli (1995) using an overlapping generations model and determining economic growth by pollution regulations and market interactions show that, depending on the decision making institution, the pollution-income relationship may have an inverted V shape, but it could also be monotonically increasing or a «sideways-mirrored S».

Another explanation relies on the fact that pollution will stop its

increase and start to decrease with economic growth because some constraints will become non-binding. Stokey (1998) shows that pollution increases linearly with income until the threshold is passed and cleaner technologies can be used. The implied pollution-income path takes the form of an inverse-V with a sharp peak, taking place at the point where a continuum of cleaner technologies becomes available. Jaeger (1998), similarly to Stokey, finds that the pollution income relationship is an inversed-V. Jaeger relies on the assumption that at low levels of pollution consumers' taste for clean air is satisfied and marginal benefit of additional environmental quality is zero.

Finally, Andreoni and Levinson (2001) suggest another explanation due to the technological link between consumption of a desired good and abatement of its undesirable byproducts (pollution). Distribution issues may be considered as not has another explanation. Torras and Boyce (1998) argue that the greater equality of incomes results in lower level of environmental degradation. This claim is challenged by Scruggs (1998).

Shafik and Bandyopadhyay (1992) estimated EKC for ten different indicators of environmental degradation (lack of clean water, ambient sulfur oxides, annual rate of deforestation, etc.). The study uses three different functional forms (log-linear, log-quadratic in income, logarithmic cubic polynomial in GDP/c and a time trend). GDP was measured in PPP and other variables included were population density, trade, electricity prices, dummies for locations, etc. Deforestation was found to be insignificant in relation to income ( $R^2$  adjusted  $\cong 0$ ).

Panayotou (1993, 1995) employed cross sectional data and GDP in nominal US \$ (1985). The equations for the pollutants considered were logarithmic quadratics in income per capita. Deforestation was estimated against a translog function in income/c and population density. All the curves estimated were inverted U's with turning point for deforestation at \$823 per capita. Panayotou used current exchange rates (instead of PPP) which lowers the income levels of developing countries compared to some developed ones.

Grossman and Krueger's (1991, 1995) and Shafik and Bandyopadhyay's (1992) suggest that at high-income levels, material use increases in a way that the EKC is N-shape. Pezzey (1989) presents arguments for an N-shape EKC and proposes that the optimal path of environmental degradation may be monotonically increasing with the level of development. However for our data this is not the case.

Cropper and Griffiths (1994) estimated three regional EKC for deforestation only. The regressions were for Africa, Latin America and Asia. They used pooled time series cross section data on a regional basis. The results for Latin America and Africa show an  $R^2$ -adjusted of 0.47 and 0.63 respectively. Both the population growth and time trend were insignificant in all areas. None of the coefficients in the Asian regression were significant and the  $R^2$ -adjusted was only 0.13. One of their main conclusions was that economic growth does not solve the problem of deforestation.

The levels of several pollutants per unit of output in specific processes have declined in the developed countries over time with the use of strict environmental regulations. Stern et al. (1996) claim that the mix of effluent has shifted from sulfur and  $\text{NO}_x$  to  $\text{CO}_2$  and solid waste, in a way that aggregate waste is still high and even if per unit output waste has declined, per capita waste may not have declined. Regressing per capita energy consumption on income and temperature gave them an inverted U-shape relationship between energy and income. Fitting a quadratic in income gave them a significant negative coefficient for the squared income term with an  $R^2$ -adjusted equal to 0.8081. Energy consumption peaked at \$14600. The authors claim that the results depend on the income measure used. If income in PPP was used the coefficient on squared income was positive but small and insignificant. If income per capita was measured using official exchange rates, the fitted energy income relationship was an inverted U-shape with squared income coefficient negative, significant (with an  $R^2$ -adjusted = 0.6564). Energy use per capita peaked at income \$23900. Table 1 presents the relevant EKC studies for sulfur.

### 3. Data

A large data set on sulfur emissions is used here (A.S.L. and Associates, 1997; Lefohn et al., 1999), which includes sulfur emissions from various fuels (hard coal, brown coal, and petroleum) as well as sulfur emissions from mining and smelting activities for most of the countries from 1850 to 1990. Emissions are based on the use of these fuels, their sulfur content, the level of smelting activity, and the sulfur retention factors. Stern and Common (2001) provide a comparison of the ASL's estimated emissions for some

developed countries. Countries like Canada, West Germany, Japan and Sweden differ substantially from the better-known OECD estimates, while the data for countries like the UK and the USA are similar. GDP per capita (in real 1990 dollars) and population data are used from the Penn World Table (Summers and Heston, 1991).

Our sample consists of the 73 countries (23 OECD and 50 non-OECD member countries), which have a full set of sulfur and purchasing power parity GDP per capita information for the period 1960-1990. The database used has 2263 observations per variable. In terms of the raw data, it is observed that emissions increase with income, but there is some sign of a decrease at high-income levels. We have used emissions rather than concentrations as the latter depend on both emissions and geographic location and atmospheric conditions in the form of wind velocity etc. We may justify the use of emissions, as there is no reason to expect that developing countries differ in any systematic manner in the dispersion of pollutants.

#### 4. Econometric methods

We have performed a Box-Cox test in order to test the linear against the logarithmic functional form of the relationship between air pollution and GDP/c. The model proposed here is a logarithmic quadratic estimated as:

$$\ln(S/c)_{it} = \alpha_i + \gamma_t + \beta_1 \ln(GDP/c)_{it} + \beta_2 (\ln(GDP/c))_{it}^2 + \varepsilon_{it} \quad (1)$$

where the  $\alpha_i$ 's are country specific intercepts and the  $\gamma_t$ 's are time specific intercepts and the countries are indexed by  $i$  and time periods by  $t$ .  $S/c$  is sulfur emissions per capita in tons of sulfur and  $\varepsilon_{it}$  is a disturbance term. Both dependent (emissions per capita) and independent (PPP GDP per capita) variables are in natural logarithms. The turning point (TP) level of income is calculated as:

$$TP = e^{\left(\frac{-\beta_1}{2\beta_2}\right)} \quad (2)$$

We have applied panel data methods to estimate the above equation. The first method employed imposes the same intercept and slope parameter for all countries and it is therefore equivalent to OLS estimation (omitted for simplicity in some cases). The second method is the fixed effects (hereafter FE) allowing each individual

**Table 1**  
**Sulfur EKC Studies**

Authors	Turning Point 1990 USD	Additional Variables	Data Source for Sulfur	Time Period	Estimation Technique	Countries/ cities
Cole <i>et al.</i> , 1997	\$8232	Country dummy, technology level	OECD	1970-92	RE, FE, OLS	11 OECD countries
Grossman and Krueger, 1991	\$4772-5965	Locational dummies, population density, trend	GEMS	1977, '82, '88	RE	Up to 52 cities in up to 32 countries
Kaufmann <i>et al.</i> , 1998	\$14730	GDP/Area, steel exports/GDP	UN	1974-89	RE, FE, OLS	13 developed and 10 developing countries
Panayotou, 1993, 1995	\$3137	-	Own estimates from fuel use data	1987-88	OLS	55 developed and developing countries
Panayotou, 1997	\$5965	Population density, policy variables	GEMS	1982-94	RE, FE	Cities in 30 develo- ped and developing countries
Seiden and Song, 1994	\$10391-10620	Population density	WRI-primarily OECD source	1979-87	RE, FE, OLS	22 OECD and 8 deve- loping countries
Shafik, 1994	\$4379	Time trend, locational dummies	GEMS	1972-88	FE	47 Cities in 31 Countries
Stern and Common (2001)	\$78703	Time and country effects	ASL	1960-90	RE, FE	74 developed and developing countries

Modified from Stern and Common (2001).

country to have a different intercept treating the  $\alpha_i$  and  $\gamma_i$  as regression parameters. This practically means that the means of each variable for each country are subtracted from the data for that country and the mean for all countries in the sample in each individual time period is also deducted from the observations from that period. Then OLS is used to estimate the regression with the transformed data. The third model is the random-effects (hereafter RE) in which the individual effects are treated as random. In this model the  $\alpha_i$  and  $\gamma_i$  are treated as components of the random disturbances. The residual from an OLS estimate of the model with a single intercept are used to construct variances utilized in a GLS estimates (for further details see Hsiao, 1986). If the effects  $\alpha_i$  and  $\gamma_i$  are correlated with the explanatory variables then the random effects model cannot be estimated consistently (Hsiao, 1986, Mundlak, 1978). Only the fixed effects model can be estimated consistently.

To control for non-observable specific effects Two Stage Least Squares (2SLS) was applied but the results were insignificant. Due to exogeneity strict requirement for the efficiency of these methods we also control for endogeneity of GDP using the GDP variable lagged two periods as the instrumental variable (IV) and in order to achieve completeness.

The orthogonality test for the RE and the independent variables is also examined. For this reason, a Hausman test is used in order to test for inconsistency in the RE estimate. This test compares the slope parameters estimated for FE and RE models. A significant difference indicates that the RE model is estimated inconsistently due to correlation between the independent variables and the error components. If there are no other statistical problems the FE model can be estimated consistently although the estimated parameters are conditional on the country and time effects in the selected sample of data (Hsiao, 1986).

We also test for serial correlation in the regression residuals, regressing the residuals on one lag of the residual and calculating the t-statistic for the autocorrelation coefficient. The Chow F tests whether pooling the data in the EU level instead of estimating separate regressions for poor and rich countries significantly reduces the goodness of fit.

We tried separating the countries according to their geographical position into Southern and Northern countries as well as according to their income level. According to the latter distinction, we have defined the first group as above average income countries and the

second as below average income countries. The turning points in this case were \$8203, \$8671 and \$7529 and \$9221 for the FE and RE and for the above- and below-average income countries.

## 5. Empirical evidence

We first present the results for the whole of the database and for the non-OECD countries as shown in Table 2 (modified and re-estimated from Stern and Common, 2001). Both the fixed and random effects models indicate the presence of a Kuznets curve and, parameter estimates as well as t-statistics are quite similar. As we observe the implied turning points are extremely high for both the fixed and random effects for the whole dataset ( $n=2263$ ) and they are equal to \$123571/c and \$91991/c respectively. Thus using the ASL database and fixed and random effect models produces a monotonic EKC for the total sample. The turning points for non-OECD countries and for the fixed and random effects models are much higher (\$501936 and \$361942 respectively).

Similarly, and from Table 3, the turning points for the OECD countries are inside the sample. Specifically for the fixed and random effects models they are equal to \$9152 and \$9166 respectively. Confirming the results derived in Stern and Common (2001), sulfur emissions per capita are a monotonic function of income per capita, when they use a global sample, and an inverted U-shaped function of income when they use a sample of high-income countries only.

The Hausman test shows that country intercepts and income are correlated in the global model. The test shows that the random effects formulation cannot be consistently estimated. This suggests that there are omitted variables which are correlated with GDP. The Chow test shows that there are differences in the estimated parameters between high- and low-income countries. The reported tests for serial correlation show that there is significant residual serial correlation in the individual countries even after common time effects have been removed<sup>1</sup>.

The F test performed in order to test for the significance of the FE shows that the null hypothesis (of non significance) is always rejected implying that the assumption of constant intercept may not always be valid for the different countries. That is, although the above-average (and the below-average) income countries exhibit



identical patterns in terms of the relationship between GDP increases and sulfur emissions, there are differences in levels as they start from different levels of sulfur emissions.

A high level of predictability is observed in both cases of FE and RE model formulation. At an income level of \$4135 (the lowest in the sample considered and for Portugal) the elasticity of emissions with respect to income was found to be 1.11. For an income level of \$5819 (Ireland) the elasticity of emissions with respect to income was found to be 0.6329, for an income of \$8838 (Austria) the elasticity is 0.049, for an income of \$9366 (Belgium) the elasticity is -0.0325 and for an income of \$11426 (Sweden) the elasticity is -0.31042<sup>2</sup>.

## 6. Abatement options for sulfur emission's reduction

Desulfurization processes exist to reduce the sulfur content of the fuel in use. The extent of removal is dependent on the physical and chemical characteristics of the sulfur in the fuel. Control technologies can be classified into three categories: 1. pre-combustion (physical coal washing and oil desulfurization); 2. during-combustion (sorbent injection and fluidized bed combustion); and 3. post-combustion (flue gas desulfurization, FGD). The choice of the technology will depend upon the characteristics of the fuel being burned and the standards for emissions, which must be met. Ease of disposition or ability to reuse waste products was found to be a secondary but important determinant of the technology used, especially as it affects the economics of certain processes.

The extent of removal is dependent on the physical and chemical characteristics of the sulfur in the fuel. Fuel cleaning techniques are relatively simple and well established but their effectiveness depends on the physical characteristics of the specific coals and crude oils, which, are subject to treatment. Fluidized Bed Combustion (FBC) can only be used for new installations and could only have an effect on total emissions over a long period. It is not possible to define abatement costs precisely since air pollution control is an integral part of the FBC boiler design. Sorbent injection could be a low cost retrofit option in cases where only moderate SO<sub>2</sub> emission reductions are required. Flue Gas Desulfurisation (FGD) is the most

**Table 2**  
Fixed and Random Effects results for the World  
and non-OECD countries

Region	World n=2263		Non-OECD n=1550	
	Fixed Effects	Random Effects	Fixed Effects	Random Effects
Constant		-24.661 (-14.029)		-19.3753 (-7.724)
ln GDP/P	4.1036 (6.141)	4.1146 (9.4749)	2.6725 (4.1386)	2.684 (4.191)
(ln GDP/P) <sup>2</sup>	-0.175 (-4.999)	-0.18 (-7.706)	-0.1018 (-2.762)	-0.10485 (-2.784)
Adjusted R <sup>2</sup>	0.144	0.155	0.143	0.151
$\rho$	0,873	0,882	0,852	0,86
AR(1)	88,3	89,7	71,59	73,71
Turning Point	123571	91991	501936	361942
Chow F Test	10.681 (0.016)	4.026 (0.04)		
Hausman Test		10.8 (P=0.0045)		13.54 (P=0.0011)

Figures in parentheses are t statistics. AR(1) is a t-test on the residual autocorrelation coefficient  $\rho$ .<sup>3</sup>

**Table 3**  
Fixed and Random Effects Results for the OECD countries

Regressors	OLS	FE	RE	TSLS
Constant	-69,711 (-11,582)		-59.59 (-18.4671)	
Ln GDP	14,261 (10,524)	12.245 (16.1613)	12.2196 (16.7735)	-1,636 (-1,382)
Ln GDP <sup>2</sup>	-0,7716 (-10,12)	-0.6712 (-16.26)	-0.6697 (-16.282)	0,1318 (1,89)
R <sup>2</sup> Adjusted	0,2998	0.311	0.332	
SE	0,565	0,29011	0,29122	0,3745
Sign. FE		2095,9		
$\rho$	0,99	0,911	0,908	
AR(1)	98,91	56,4	54,94	
Turning Point	10313	9152	9166	
Hausman test			0.25	

commercially developed technology and the only one available for achieving very high removal efficiency at all types of installation, new or retrofit. The general trend is for Sorbent Injection (SI) to have the lowest capital costs among pre-combustion technologies, FBC and spray-dry scrubbers next, followed by wet scrubbers with regenerable processes having the highest capital costs.

Cost estimates for each technology are influenced by fuel type, plant size, sulfur content of the fuel, new or retrofit application, labor, construction and electricity cost factors. The slopes of the total abatement cost curves differ from country to country and if, for any given abatement level, the slope of the total abatement cost curve for one country is steeper than for another, then the abatement cost in the first country is higher than in the second. Given projections of uncontrolled emissions, estimates can be made of the potential for their reduction using available abatement technologies and of the likely cost. Table 4 presents information on the cost-effectiveness of the available technologies as well as the applicability, the capital, operating and maintenance costs.

The selection of appropriate strategies to reach and implement pollution control objectives is of crucial importance to planners. Because of the existing differences between countries in energy-use patterns, emissions, source locations and other economic factors, it is unlikely that a single uniform program of secondary abatement will be appropriate in all countries. To reduce sulfur emissions, the national decision maker may set a maximum allow-able rate of pollution output for each generic type of source (electricity generating, industry, petroleum refineries and transport) by type of pollutant. Furthermore, fuel quality regulations can be structured around the types of fuels in use (e.g. coal, oil etc) and can be limited by the technical possibilities and the costs of cleaning process for the different fuels.

Reducing energy consumption through either conservation or energy improvements can also reduce sulfur emissions. The latter can be achieved for instance by reducing energy consumption through more efficient generation, use of combined heat and power, etc. Denmark, Norway and Sweden are the only OECD countries that continue to increase energy taxation since the 1980s aiming to encourage energy conservation. On the other hand, IEA claims that inconsistency of taxation and energy policy is evident in the UK, which actually discourages efficiency by charging VAT on home-insulation products. IEA suggests that the potential for demand reductions for

the UK, the Netherlands, Austria, the EEC and Western Europe in the industrial sector in 2000 can be as high as 25%, 21%, 10%, 25% and 30% respectively. At the same time Sweden can achieve demand reductions of 50% in its residential sector and 40% in its commercial sector. The EEC countries and the Netherlands can achieve an average of approximately 30% and 21% savings through cost-effective means in the residential and commercial sectors respectively. A 30% increase in energy efficiency can reduce energy requirements by 25%, which is equivalent to more than 1200 million tons of oil equivalent per year in 2000 (IEA, 1987).

Low sulfur coal may be a good way to reduce emissions where emission standards are met by using coal within a specific range of sulfur content. For instance, a standard of 2000 mg/m<sup>3</sup> is equivalent to approximately 1% sulfur content of coal, as the cut-off level above which sulfur abatement technologies would be used. Emission standards between 1000 and 2000 mg/m<sup>3</sup> are equivalent to coal sulfur content of 0.5-1% and there is no percentage removal requirement. Plants facing these standards can use either low sulfur content coal alone, or in conjunction with a limited-efficiency abatement technique (Vernon, 1989).

The use of low sulfur coal is a function of its availability and its cost relative to other control methods. Obviously, if the demand for low sulfur coal increases then both price and availability will change. Technical barriers exist to using low sulfur coal because it has a low calorific value and different ash characteristics which affect the operation of electrostatic precipitators. There are political barriers when no local supplies exist and government energy policies restrict, through import quotas, the import of supplies from elsewhere (Germany, Spain).

Substitution of fossil fuels by nuclear power and natural gas is also possible. But nuclear and hydropower have seen opposition on environmental grounds while other non-fossil fuel sources have been under-developed. Public pressure may increase the demand for gas fired power plants. High capital cost and costs of decommissioning mean that the nuclear plants have no advantage over coal-fired plants with secondary emissions control. The costs of NO<sub>x</sub> and SO<sub>2</sub> controls on coal-fired plants are similar to those of gas firing plants. A range of \$36-\$50 m per kWh (US \$1987) for coal-fired plants with full environmental control compares with \$44-\$48 m per kWh for a natural gas plant meeting similar standards (assuming a coal price of \$40-60 per ton and a relative gas price of \$160 per ton

**Table 4**

Sulfur emission abatement options and costs (in \$ million 1985).

Costs are based on a new 500 MW power plant,  
using hard coal with 2% sulfur content, 70% load factor  
and 5% retention factor

Abatement Method	Applicability	Sulfur removal efficiency (%)	Capital Cost	Operating and Maintenance cost		Cost-Effectiveness
				FIXED	VAR	\$/t SR
Fuel switching (e.g. oil to gas)	All Users	Up to 99	-	-	-	(1)
Physical coal cleaning	All users	25	-	-	-	635-1625
Heavy fuel oil desulfurisation	All users	80	7.775	6.32	12.28	2100-2930
Sorbent injection	All users	50	0.344	0.22	2.59	485-750
Atmospheric Fluidised Bed Combustion	Power plants, industrial boilers	80	3.259	0.16	2.71	238-446
Circulating Fluidized Bed Combustion	New plants only	85	7.061	0.35	4.77	529-835
Flue Gas Desulfurization	Power plants, industrial boilers and process emissions	90	29.462	1.67	4.02	650-1200
Gas Oil Desulphurization	All users	90	1.918	1.93	2.2	2900-3740

(1) Depends on relative price and sulfur content

Source: Halkos (1995)

coal equivalent) (IEA, 1988). Newbery (1993) cites that if FGD investment is coordinated with gas, then capital costs of a gas burning Combined Cycle Gas Turbine (CCGT) will be £360-£500/kW compared with FGD capital costs of £150-£175/kW capacity. According to the same source, at the 1993 import parity price of coal, FGD was cheaper than new CCGT stations. This implies that FGD can be competitive with CCGT.

**Table 5**  
Technologies applied by fuel used

Existing plants	<b>HARD COAL</b> Hard Coal Washing (HCW) Sorbent Injection (SI) Flue Gas Desulfurization (FGD) Combination of HCW and SI Combination of HCW and FGD
New plant (less or equal to 500 MWe)	All the above technologies and additionally: Atmospheric Fluidized Bed Combustion (AFBC) Circulating Fluidized Bed Combustion (CFBC) Combination of HCW and AFBC Combination of HCW and CFBC
Existing plants	<b>BROWN COAL</b> Sorbent Injection (SI) Flue Gas Desulphurization (FGD)
New plants	All the above technologies and additionally: Atmospheric Fluidized Bed Combustion (AFBC) Circulating Fluidized Bed Combustion (CFBC)
	<b>HEAVY FUEL OIL</b> Heavy Fuel Oil Desulfurization (HFOD) Flue Gas Desulfurization (FGD) Combination of HFOD and FGD
	<b>GAS DIESEL OIL</b> Gas Diesel Oil Desulfurization (GDOD) Flue Gas Desulfurization (FGD) Combination of GDOD and FGD
	<b>PEAT AND BROWN COAL BRIQUETTES</b> Sorbent Injection (SI) Flue Gas Desulfurization (FGD)

Source: Halkos (1995)

But would a massive program of sulfur control have a large effect on the prices which consumers pay for electricity? Highton and Webb (1984) showed that the increased price for the consumer would be about 4% in electricity costs with a 50% reduction in Central Electricity Generating Board of England and Wales (CEGB) emissions of sulfur. For large industrial consumers the effect would be an increase of slightly over 5%. Of course, this percentage increase in the electricity cost will vary from country to country due to different domestic unrestricted tariffs or different industrial tariffs.

The cost of an emission abatement option is given by the total annualized cost (TAC) of an abatement option, including capital and operating cost components:

$$TAC = [(TCC) * (r / (1-(1+r)^{-n}))] VOMC + FOMC$$

Where TCC is the total capital cost (\$), VOMC and FOMC are the variable and fixed operating and maintenance costs (\$) respectively and  $(r/(1-(1+r)^{-n}))$  is the capital recovery factor at real discount rate  $r$ , which converts a capital cost to an equivalent stream of equal annual future payments, considering the time value of money (represented by the discount rate,  $r$ );  $n$  represents the economic life of asset (in years). The estimation of the annual operating and maintenance costs requires a great deal of information (for example, the sulfur content of fuel used, the annual operating hours, removal efficiencies of the control methods, etc). It consists of a fixed portion that is dependent on the use of the plant (e.g. maintenance and labor costs) and a variable portion dependent on the prices for electricity, labor, sorbents and waste disposal and the specific demand for energy due to the abatement process. Table 5 presents the control technologies applied to each fuel type while table 6 presents the fuels used by each sector and to which we apply the available abatement technologies.

## 7. Conclusions and policy implications

Like inequality, pollution tends to become worse before it becomes better along a country's development path. Our results indicate the existence of an inverted U-shaped relationship between economic development and pollution in the form of sulfur emissions as shown in Figure 1. The turning point occurs at \$9152 for the OECD countries, at \$501936 for the non-OECD countries and at \$123571 for the world in general.

Specifically, using this panel database and fixed and random effect models produces a monotonic EKC for global and non OECD samples with extremely high turning points and an inverted U-shaped curve within the sample turning points for the case of OECD countries. Estimating an EKC using data for only the OECD countries leads to estimates where the turning point is not biased downwards relative to those estimated using data for the World as a whole.

**Table 6**  
**Fuels in which control technologies**  
**are applied in each sector**

Sectors	Fuels	Hard Coal	Brown Coal	Brown Coal Briquettes	Heavy Fuel Oil	Gas Diesel Oil	Peat	Refinery Fuel Oil
Electricity Generating		*	*	*	*	*	*	
Industry Energy		*	*	*	*	*	*	
Transport		*			*	*		*
Others		*			*	*		

Source: Halkos (1995)

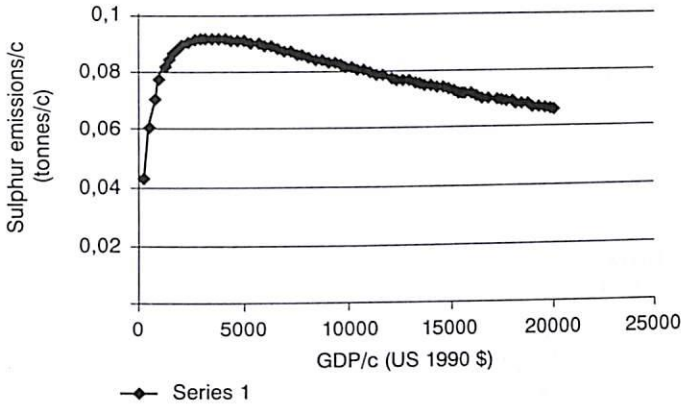
The acceptance of an EKC hypothesis means that there is an inevitable level of environmental damage that follows a country's development at the earlier stage but with a significant improvement at a later stage of this country's economic growth. Thus, an EKC is the result of structural change that follows economic growth. However, this may not be optimal if environmental critical loads are crossed irreversibly. The positively sloped part of an EKC where growth is worse may take a long time to cross. This implies a present value of higher future growth and cleaner future environment may be offset by high current rates of environmental damage. At the same time it may be cheaper to abate today than in the future.

The decomposition of the EKC into its main determinants shows that economic growth increases pollution levels due to scale and industrialization but ignores the abatement effect of richer countries (Panayotou, 1997).

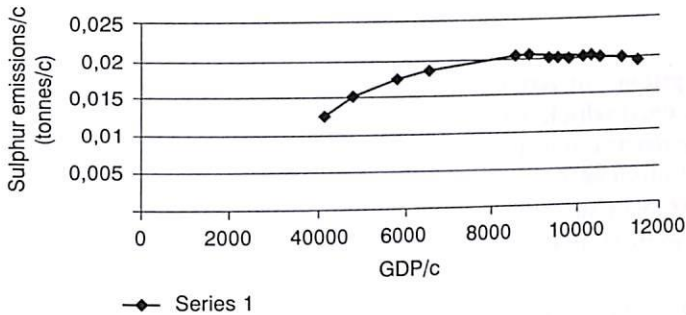
Acceptance of an EKC may seem as a temporary phenomenon and we may seek ways to stimulate growth like trade liberalization, price reform, economic restructuring, etc. Some of the steepness of an inverted U-shaped relationship between environmental damage in the form of pollution and economic growth is caused by various policy distortions such as protection of industry, energy subsidies, etc. Developing countries can flatten out their EKC's by defining and applying property rights over natural resources, eliminating any policy distortions and internalizing environmental costs to the sources that generate them (Panayotou, 1993). Additionally, the improper allocation of property rights may lead to market failure.



**Figure 1**  
A global EKC between S/c and GDP/c



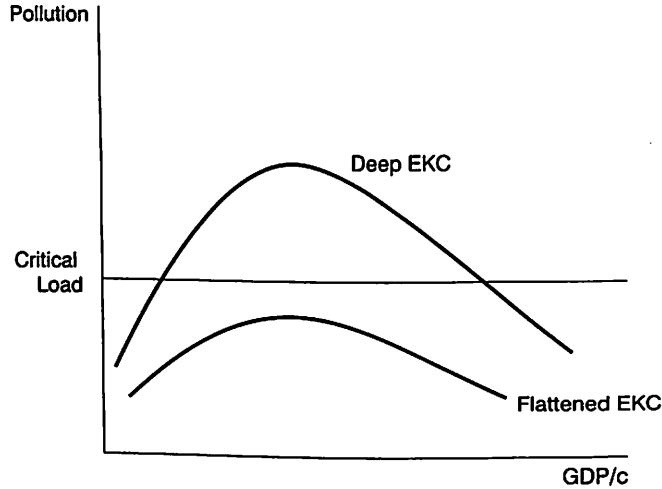
**Figure 2**  
An EU EKC between S/c and GDP/c



A kind of development assistance can be organized in order to flatten out developing countries' EKC by making environmental protection an integral part of the OECD's financial policy. In the case that ecological thresholds, in the form of critical loads that might be crossed irreversibly, then a deep EKC implying high rates of pollution per unit of GDP/c increase is not optimal economically and environmentally as more of both can be attained with a better managed use of the same resources.

A part of the steepness of the inverted U-shaped relationship between economic growth and pollution is due to policy distortions

**Figure 3**  
EKC and critical loads



Source: Modified from Panayotou (1993)

(under-pricing of natural resources, subsidies of energy and agrochemicals, etc), which are at the same time environmentally and economically destructive. Governments can flatten out their EKC by reducing or eliminating policy distortions, defining and applying property rights over natural resources and internalizing environmental costs to the sources that generate them.

The need for technology transfer to help developing countries to achieve sustainability emerges. The main idea is that abatement technologies in developed countries are cleaner and more advanced. As developing countries have no financial resources to import and use these technologies at commercial cost, this implies that developed countries should transfer or facilitate the transfer of these technologies to less developed or developing countries. The impact of this technology transfer depends on the type of industrial activity. That is, in the energy sector these transfers will be more beneficial for the environment compared to other industries such as textiles, etc. It should be emphasized that transfer of information must accompany these technology transfers on know-how and skills to enable countries to design or modify their own technologies.

## Notes

1. A model estimated in first differences reduces statistical problems but results in a monotonic EKC when estimated on both high and low income samples. Stern and Common (2001) provide the results in first differences where the turning points again differ substantially. They equal \$53590, \$586965 and \$21545 for the global, OECD and non-OECD samples respectively.

2. GDP may be an integrated variable at least in the case of the Western European countries (Stern, 1998; Perman and Stern, 1999). The Hausman tests reported in Table 2 show that there may be omitted variables correlated with GDP. If the EKC regressions do not co-integrate the estimates may be spurious and non-co-integration is very possible. The very high reported autocorrelation coefficients in Stern and Common (2001) imply this conclusion. Thus the regression results reported above may be spurious. Differencing the data will eliminate potential stochastic trends in the series.

3. Data are for the following countries:

OECD: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Greece, Ireland, Italy, Japan, Luxembourg, Netherlands, Norway, New Zealand, Portugal, Spain, Sweden, Switzerland, Turkey, UK, USA, West. Germany.

Non-OECD: Algeria, Argentina, Barbados, Bolivia, Brazil, Chile, China, Colombia, Cyprus, Czechoslovakia, Egypt, Ghana, Guatemala, Honduras, Hong Kong, India, Indonesia, Iran, Israel, Kenya, Korea, Madagascar, Malaysia, Mexico, Morocco, Mozambique, Myanmar, Namibia, Nicaragua, Nigeria, Peru, Philippines, Romania, South Africa, Saudi Arabia, Singapore, Sri Lanka, Syria, Taiwan, Tanzania, Thailand, Trinidad & Tobago, Tunisia, Uruguay, U.S.S.R., Venezuela, Yugoslavia, Zaire, Zambia, Zimbabwe.

## References

- Andreoni J. and Levinson A., «The Simple Analytics of the Environmental Kuznets Curve», *Journal of Public Economics*, 80, 2001, 269-286.
- Ansuategi A., Barbier E.B. and Perrings C.A., The Environmental Kuznets Curve, in: Van den Bergh, J.C.J.M. and Hofkes, M.W. (Eds) *Theory and Implementation of Economic Models for Sustainable Development*, Kluwer, 1998.
- Arrow K., Bolin B., Costanza R., Dasgupta P., Folke C., Holling C.S., Jansson B-O, Levin S., Maler K-G, Perrings C. and Pimentel D., *Economic Growth, Carrying Capacity and the Environment*, Science, 268, 1995, 520-521.

- A.S.L. and Associates, *Sulphur Emissions By Country And Year*, Report No: DE96014790, US Department of Energy, Washington DC, 1997.
- Cole M.A., Rayner A.J. and Bates J.M., The Environmental Kuznets Curve: An Empirical Analysis, *Environment and Development* 2 (4), 1997, 401-416.
- Cropper M. and C. Griffiths, The Interaction of Pollution Growth and Environmental Quality, *American Economic Review*, 84(2), 1994, 250-254.
- de Bruyn, S.M., Explaining the Environmental Kuznets Curve: Structural Change and International Agreements in Reducing Sulphur Emissions, *Environment and Development Economics*, 2(4), (1997), 485-503.
- Dinda S., D. Coondos and M. Pal, «Air Quality and Economic Growth: An Empirical Study», *Ecological Economics*, 34, 2000, 409-423.
- Ekins P., The Kuznets Curve for the Environment and Economic Growth: Examining the Evidence, *Environment and Planning A* 29, 1997, 805-830.
- Grossman G. and A. Krueger, Environmental impacts of a North American Free Trade Agreement, NBER Working Paper 3914, *National Bureau of Economic Research*, Cambridge MA., 1991.
- Grossman G. and A. Krueger, Economic Growth and the Environment, *Quarterly Journal of Economics*, 110, 1995, 353-377.
- Halkos G., An Evaluation of the Direct Cost of Abatement Under the Main Desulfurization Technologies, *Energy Sources*, 17 (4), 1995, 391-412.
- Halkos, G. and E. Tsionas, Environmental Kuznets Curve: Bayesian Evidence from Switching Regime Models, *Energy Economics*, 23 (2), March 2001, 191-209.
- Han X. and Chatterjee L., «Impacts of Growth and Structural Change on CO<sub>2</sub> Emissions of Developing Countries», *World Development*, 25, 1997, 395-407.
- Highton N.H. and Webb, M.G., Sulphur Dioxide from Electricity Generation: Policy Options for Pollution Control, *Energy Policy*, March 1980, pp. 61-76.
- Highton N.H. and Webb M.G., The Effects on Electricity Prices in England and Wales of National Sulphur Dioxide Emission Standards for Power Stations, *Journal of Environmental Economics and Management*, 11, 1984, 70-83.
- Hsiao C., *Analysis of Panel Data*, Cambridge University Press, Cambridge UK, 1986.
- IEA, *Energy Conservation in IEA Countries*, International Energy Agency, Paris France, OECD/IEA, pp 68-70, 1987.
- IEA, *Energy Policies and Programmes of IEA Countries*, 1987 Review,

- International Energy Agency 494 pp, Paris France: OECD/IEA, 1988.
- Jaeger W., *A Theoretical Basis for the Environmental Inverted-U Curve and Implications for International Trade*, Williams College, Working Paper, 1998.
- Jones L.E. and Manuelli R.E., A Positive Model of Growth and Pollution Controls, NBER, *Working Paper* 5205, 1995.
- Kaufmann R.K., Davidsottir B., Garnham S. and Pauly P., The Determinants of Atmospheric SO<sub>2</sub> Concentrations: Reconsidering the Environmental Kuznets Curve, *Ecological Economics*, 25, 1998, 209-220.
- Kuznet S., *Economic Growth and Structural Change*, New York, 1965.
- Kuznet S., *Modern Economic Growth*, Yale University Press, 1966.
- Lefohn A. S., D.H. Janja and B.H. Rudolf, «Estimating Historical Anthropogenic Global Sulphur Emission Patterns for the Period 1860-1990», *Atmospheric Environment*, 33, 1999, 2435-2444.
- Lopez R., The Environment as a Factor of Production: The Effects of Economic Growth and Trade Liberalization, *Journal of Environmental Economics and Management*, 27, 1994, 163-184.
- McConnell K.E., «Income and Demand for Environmental Quality», *Environment and Development Economics*, 2, 1997, 383-399.
- Mundlak Y., On the Pooling of Time Series and Cross Section Data, *Econometrica*, 46, 1978, 69-85.
- Newbery D., The Impact of EC Environmental Policy on British Coal, 28 pp. *Oxford Review of Economic Policy*, 11, 1993, 297-346.
- Panayotou T., Empirical Tests and Policy Analysis of Environmental Degradation at Different Stages of Economic Development, Working Paper WP238, Technology and Employment Programme, *International Labor Office*, Geneva, 1993.
- Panayotou T., Environmental Degradation at Different Stages of Economic Development, in: I. Ahmed and J. A. Doeleman (eds.) *Beyond Rio: The Environmental Crisis and Sustainable Livelihoods in the Third World*, MacMillan, London, 1995.
- Panayotou T., «Demystifying the Environmental Kuznets Curve: Turning a Black Box into a Policy Tool», *Environment and Development Economics*, 2(4), 1997, 465-484.
- Pezzey J.C.V., Economic Analysis of Sustainable Growth and Sustainable Development, Environment Department Working Paper No. 15, *The World Bank*, Washington D.C., 1989.
- Perman R. and Stern D.I., *The Environmental Kuznets Curve: Implications of Non-Stationarity*, Working Papers in Ecological Economics 9901, Center for Resource and Environmental Studies, Australian National University, 1999.
- Rock M.T., «Pollution Intensity of GDP and Trade Policy: Can the World Bank Be Wrong?», *World Development*, 24, 1996, 471-479.

- Scruggs L.A., «Political and Economic Inequality and the Environment», *Ecological Economics*, 26, 1998, 259-275.
- Selten T. and D. Song, Environmental Quality and Development: Is There a Kuznets Curve for Air Pollution Emissions? *Journal of Environmental Economics and Management*, 27, 1994, 147-162.
- Shafik N. and S. Bandyopadhyay, Economic Growth and Environmental Quality: Time Series and Cross-Country Evidence, Background paper for the World Development Report 1992, *The World Bank*, Washington DC, 1992.
- Shafik N., Economic Development and Environmental Quality: An Econometric Analysis, *Oxford Economic Papers* 46, 1994, 757-773.
- Rock M.T., Pollution Intensity of GDP and Trade Policy: Can the World Bank Be Wrong?, *World Development*, 24, 1996, 471-479.
- Stern D. I., M.S. Common and E.B. Barbier, Economic Growth and the Environmental Degradation: The environmental Kuznets Curve and Sustainable Development, *World Development*, 24, 1996, 1151-1160.
- Stern D.I., Progress on the Environmental Kuznets Curve? *Environment and Development* 3, 1998, 175-198.
- Stern D. I. and M.S. Common, «Is there an Environmental Kuznets Curve for sulphur?», *Journal of Environmental Economics and Management*, 41, 2001, 162-178.
- Stokey N.L., «Are There Limits to Growth?», *International Economic Review*, 39, 1998, 1-31.
- Suri V. and Chapman D., 'Economic Growth, Trade and Energy: Implications for the Environmental Kuznets Curve, *Ecological Economics*, 25, 1998, 195-208.
- Summers R. and A. Heston, «The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988», *Quarterly Journal of Economics*, 106, 1991, 1-41.
- Torras M. and J.K. Boyce, «Income, Inequality and Pollution: A Reassessment of the Environmental Kuznets Curve», *Ecological Economics*, 25, 1998, 147-160.
- Vernon J.L., *Market Impacts of Sulphur Control: The Consequences for Coal*, IEA Coal Research, October 1989, London, 1989.

